DOCUMENT RESUME

ED 031 313 24 PS 002 102

By-Bommarito, James W.
Verbal Reinforcement as an Adjustment Predictor with Kindergarten Children. Final Report.
Illinois State Univ., Normal.
Spons Agency-Office of Education (DHEW), Washington, D.C. Bureau of Research.
Bureau No-BR-7-E-098
Pub Date Dec 68
Grant-OEG-0-8-0098-1910-010
Note-136p.
EDRS Price MF-$0.75 HC-$6.90
Descriptors-Discrimination Learning, *Emotional Adjustment, *Emotionally Disturbed Children, Kindergarten Children, Negative Reinforcement, Personality Tests, *Predictive Ability (Testing), *Screening Tests, Socialization, Socioeconomic Status, Task Performance, *Verbal Operant Conditioning
Identifiers-Peabody Picture Vocabulary Test

In an attempt to develop an instrument to screen emotionally disturbed children, a study was designed to verify and extend previous research findings. These findings indicated that adjusted children in kindergarten exhibited greater conditionality on a simple discrimination learning task than maladjusted pupils under a process of continuous mild verbal punishment for every undesirable response. Of the 224 children randomly selected with stratification for sex, 30 were chosen for each group (adjusted and maladjusted) on the basis of extreme scores on three personality scales, the Problem Checklist, the Behavior Checklist, and the Minnesota Scale. The experimental task itself involved a choice between a picture of a human and a picture of a toy with a verbal "that's bad" from the examiner contingent on each toy choice. A subject's score was his increase in human choices from his first to his fourth block of 25 trials. The influence of several intervening variables was evaluated: verbal reinforcement as an effective punishment (significant and cummulative in effect); correlation of mental age, as reasured by the Peabody Picture Vocabulary Test, to task score (insignificant); examiner effects (insigificant); and correlation of task score to socioeconomic status (insignificant). A reevaluation of the data after the study partially supported the main hypothesis. (MH)

ED031313

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

BR 7-E-098
PA 24
OE/BR

FINAL REPORT

Project No. 7-E-098
Grant No. OEG-0-8-0098-1910 (010)

VERBAL REINFORCEMENT AS AN ADJUSTMENT PREDICTOR WITH KINDERGARTEN CHILDREN

James W. Bommarito

December, 1968

The research herein was performed pursuant to a grant with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

Illinois State University

Normal, Illinois

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

PS002102

TABLE OF CONTENTS

Acknowledgements iii

Summary 1

I. Introduction 3

II. Some Related Crucial Issues 9

III. Hypotheses and Sampling Procedure 18

IV. Method and Instrumentation 22

V. Results 28

VI. Discussion 42

VII. Concluding Comments 65

References 70

Appendixes 77

- A. Problem Checklist 78
- B. Behavior Checklist 85
- C. Minnesota Scale 89
- D. Socialization Scale 93
- E. School Observation Schedule 99
- F. Step by Step Report of Research Progress 102
- G. Forms for Recording Data from Learning Task and Peabody 106
- H. Approval Letter to School Personnel 109
- I. Letter from School Officials to Parents 111
- J. General Directions for Kindergarten Teachers 113
- K. Details on Maximization of Block Effects 115

List of Tables

Table I. Number of Alternates by Sex and Conditions of Testing 21

Table II. Summary for the Analysis of Variance on the entire Sample over the Three Blocks (1964 and 1968 Samples Compared) 118

Table III. Significant Differences in Weighted Mean Gain Scores Between Blocks of Reinforcement for all Possible Paired Comparisons in 1968 Sample 29

Table IV. Significant Differences in Weighted Mean Gain Scores Between Blocks of Reinforcement for all Possible Paired Comparisons in the Entire Sample (1964 and 1968 Samples Compared) 119

Table V. Summary on a Two-Way Analysis of Variance for Gain Scores by Children with Four Classifications of Adjustment, Six Conditions of Reinforcement, and Interaction (1964 and 1968) 120

Table VI. Raw Scores for Two-Way Analysis of Variance (1964) 121

Table VII. Raw Scores for Two-Way Analysis of Variance (1968) 123

Table VIII. Means and Standard Deviations of Gain Scores (1964 and 1968) 125

Table IX. Comparison of Mean Weighted Gain Scores Between Social Classes in Descending Order of Differences 127

Table X. General Summary of Significant Statistics (1964 and 1968) 128

Table XI. Summary of Differences in Weighted Mean Gain Scores in Block 4 Under Each Examiner Within Subgroups of Each Adjustment Category (1968) 131

Table XII. Relationship Between Extremes in Gain Scores and Adjustment Status (Fisher's Exact Probability Test) 132

ACKNOWLEDGMENTS

This study has been dependent upon the skills, enthusiasm, and cooperation of many educators over a one-year period of operation.

Without the cooperation of Dr. Fred McDavid, Superintendent of Schools in Bloomington, in making the services and facilities of the school system available, the study could not have been undertaken. Mr. Don Henderson, the Assistant Superintendent in Charge of Curriculum Development, was especially helpful in expediting matters necessary for the progress of the research. The splendid cooperation of the principals in the elementary school was also highly appreciated. The investigator, however, is especially grateful to the participating teachers for the enthusiasm and proficiency with which they performed their responsibilities.

Finally, Dr. Richard Reiter and his staff of the computor services at Illinois State University provided the calculations on all the parametric statistics of this study. The author is certainly quite appreciative of this considerable aid.

SUMMARY

The present study attempted to verify and extend findings from previous research which indicated that adjusted children in kindergarten exhibited greater conditionality on a simple discrimination learning task than maladjusted pupils under a process of continuous mild verbal punishment for every undesirable response. The social-learning theory underlying this expectation was explained in full. Discussion also centered on possible advantages of the conditioning task as a screening instrument for emotional disturbance. The value and need of such an instrument was elaborated, especially in view of (1) the increasing attention to the education of the emotionally disturbed child, (2) the vital role of the screening process itself in this education, and (3) the present unavailability of any satisfactory compact procedure of this type.

In the research design itself, randomly selected kindergarten children, with strata for sex, were rated on three personality scales which had statistical support not only on the basis of reliability but also on the basis of internal consistency, concurrent validity, and predictive validity. The children who scored at or above P_{73} or its equivalent on all three scales were designated as maladjusted; those who scored at or below the P_{27} level or its equivalent on all three scales were classified as adjusted. Children in both classifications were randomly assigned to two examiners with consideration of the sex stratum. Each examiner administered the Peabody Picture Vocabulary Scale and a Discrimination Learning Task to every child in his assigned group. The latter task consisted of 25 cards with two objects on each card: a toy and a human. Each child was presented with four blocks of the cards, a block being the entire set of 25 cards. On the first block, the examiner simply requested the child to give his object preference. The number of humans thus selected constituted his base score. Beginning with the second block and continuing through the fourth block, the examiner said "that's bad" every time the child chose a toy. The criterion measure was the frequency of human preferences. In the actual scoring of human preferences, the base score was subtracted from succeeding block scores and a constant of 25 then added. This process was followed to eliminate possible initial variability. Analysis of variance and _t_ tests were based on these "gain scores" for Blocks 2, 3, and 4. As part of this analysis, the principal hypothesis of the study was tested; namely, that conditioning by mild verbal punishment would show differential gain scores between adjusted and maladjusted pupils. In the original study, the null hypothesis was rejected (with $p < .05$) by the use of parametric statistics. In the present replication, the null hypothesis was also rejected with the application of non-parametric statistics after adjustments for possible procedural errors. The data, in effect, provide only partial verification of the original findings and suggest the need for additional cross-validation. Suggestions for this purpose have been outlined.

Statistical analyses were also applied for possible differences in gain scores due to the variables of examiners, intelligence, sex, amount of reinforcement, and social class status. In areas of agreement between the original study and the replication, the most significant findings were (1) that no significant sex differences emerged, (2) that intellectual level was unrelated to the criterion measure; (3) that extremely significant increments in gain scores were due to the amount of reinforcement over blocks, with maximum effects at the end of three blocks; and (4) that interaction effects between examiner and the adjustment status of the child were insignificant. But the results from the two studies varied with respect to the variables of examiners and social class status. In the earlier research, examiner differences were found in the maladjusted group. In the replication, examiner differences were noted only for "culturally deprived" children. Neither study indicated differential gain scores in paired comparisons between children from any two social classes. The original study, however, did show a statistically significant relationship between type of paternal occupation and gain scores of children, a finding which was not verified in the replication. In the cross-validation research, adjusted children attained statistically significant greater scores on socioeconomic indices than the maladjusted pupils.

The implications of these results and differences from the two studies were discussed in terms of methodological considerations, challengable premises, and alternative approaches in future research with more potential for a functional analysis of behavior.

In an attempt to account for the differential responsiveness of children with variations in adjustment status, several factors were discussed. These were (1) the basis for the effectiveness of mild verbal punishment in general, (2) some implications regarding its use in the socialization process, and (3) the reasons for its discriminatory function with respect to adjustment status. With regard to the effectiveness of negative reinforcement in predicting adjustment status, emphasis was placed on the development of rewards and punishments in the reinforcement life histories of the individuals as generalized reinforcers.

As a guide for future research, some attention was also devoted to the bases for confusion in the literature on verbal conditioning. Finally, possible directions for future research were discussed. These related to immediate needs for the development of the learning task as a screening instrument, numerous potential practical uses of the learning task, and further extensions and applications.

I. INTRODUCTION

At present, the importance of selected programming for emotionally disturbed children is widely recognized throughout the United States. "Schools have been pressed . . . to initiate programs for the promotion of mental health." (McNabola, 1961, p. 82). A recent publication in the education of the emotionally disturbed child has emphasized: "The . . . teacher's number one problem has been, is now, and probably will continue to be, the need for mental health consultation in conjunction with special school programs for the emotionally disturbed children in their classrooms." As an indication of the "shocking and overwhelming" enormity of the problem, the authors cited the following trends: (1) Approximately only 2% of the nation's "severely disturbed" children are receiving psychotherapy; (2) one out of every five school children has an identifiable learning and/or emotional disability; (3) the number of universities offering teacher training courses on the education of the emotionally disturbed child increased from five (1955) to sixty-eight (summer, 1965); special classroom enrollment for emotionally disturbed children on a national basis doubled from 1948 to 1958 (Long, Morse, and Newman, 1965, p. vii). Other indications stress the importance of this programming. These relate to the need for a preliminary taxonomy of procedures involved in conducting special classes (Hollister and Goldston, 1962); to the variety of programs some states have developed for this group (Bower, et. al., 1961) and to the fact that some states such as Illinois (House Bill 1407) and Minnesota (Balow, 1966) have felt it necessary to develop mandatory legislation for special classes in the area of emotional disturbance.

In connection with this programming, the role of screening for emotional disturbance has been deemed crucial. Thus, to Bower, the importance of early identification and, hence, early screening rests on three widely accepted premises:

> "1. The earlier a developing problem can be detected, the more economically and effectively it can be corrected.
>
> 2. The younger the child is, the more amenable to change are his personality and outlook on life.
>
> 3. Personality is formed chiefly by the interpersonal relationship of a child with his adults and peers" (Bower and Lambert, 1961, p. 9).

Despite the need for early screening, teachers sometimes lack the training and systematic theoretical framework which would enable them to detect early signs of emotional disturbance. The crucial role of screening and the inadequacies of teachers in performing this task greatly emphasize the need for systematic early identification. At the present time, the instruments Bower and his colleagues have developed for this purpose probably continue to be unsurpassed. At present, these remain research tools (Bower and Lambert, 1961).

Their validity, however, appears to rest principally on the agreement between teacher ratings and the judgement of clinical workers (Bower, 1958; Bower and Lambert, 1961). But White and Harris have severely criticized this basis of validity on several grounds. First, the data makes it unclear whether or not teachers were aware of children who had been examined by mental health personnel. Such knowledge conceivably could have contaminated the teachers' ratings of adjustment. Secondly, as these authors emphasized, the emotionally disturbed children were so judged by the mental health personnel using unstated criteria. As these writers further stressed, the unbiased nature of the ratings and the definition of emotional disturbance constitute crucial elements for studies, such as Bower has conducted (White and Harris, 1961, p. 171.)

Bower's instruments also apparently present two other difficulties: First, their scoring and administration seemingly involve considerable time for the teacher, particularly at the primary level for the administration of the tests and the scoring and classification of the responses (Bower and Lambert, 1961a). Secondly, they require that the teachers observe the children from two to five months before rating them (Bower and Lambert, 1961a, p. ix). Hence, they are conceivably not maximally effective as screening instruments in that they do not identify the child at the beginning stage before he has had time to develop maladaptive behavior. In view of these limitations, the many conditions posited for an effective screening instrument (Bower and Lambert, 1961a, pp. 3-14), and the importance of false negatives and positives (Anderson, 1960, p. 4), the conclusion by Anderson and others that a good compact screening instrument has yet to be developed and will be extremely difficult to develop is easily understandable and should be readily accepted to research workers (Anderson, 1959, p. 4; Anderson, 1960, p. 70).

In a previous study by Bommarito (1964), he hypothesized that susceptibility to continuous mild verbal punishment would differentiate between adjusted and maladjusted kindergarten children who were so designated by extreme cut-off scores on three teacher rating scales. In short, the hypothesis was that response to verbal operant conditioning would serve as a screening instrument for emotional disturbance. In order to appreciate this theoretical position, the related literature needs to be discussed briefly.

Recently a number of research studies and theoretical positions by learning theorists have emphasized that socialization occurs primarily as a function of the reward characteristics of parents and others, coupled with anxiety about the loss of this nurturance (Bandura, 1959; Bronfenbrenner, 1961; Mussen, 1960; Mussen and Conger, 1956; Sears et al., 1957; and Wright, 1957). McCandless (1961, pp. 325-26) has summarized these theoretical foundations in all their essentials.

"To become socialized...a child must have learned that his parents and (through generalization) other people are important and that he is dependent on them. Dependency, as a *theoretical construct*, should be distinguished from dependent behavior. In the first sense, it refers to the *reward characteristics of parents and others*, which have developed through a combination of good mothering and fathering, coupled with anxiety about the loss of this parental nurturance. In the second and more common sense, it refers to the clinging, lap-sitting, affection-and assistance-seeking behavior usually referred to as dependent.

These parents who have helped their child to acquire dependency (in the theoretical sense of the word) have provided a powerful tool for his later, more complicated socialization. If his parents, that is, are generally important and generally regarding to him, then, through the process of [stimulus] generalization, other adults, children, and their rules also become important."

This theoretical position regarding the nature and value of early socialization has also received support from empirical studies. In this connection, Radke has studied the child-rearing practices utilized with pre-school children (1946). Similarly, Sears and his workers have investigated the same area with respect to kindergarten children (1957). Though both workers have found that verbal prohibitions and restrictions are used frequently in numerous aspects of the socialization process, the logical inference would be that this verbal punishment would be ineffective in the absence of this positive reinforcement history. Among others, Sears (1957) and Bronfenbrenner (1961) have supported this view. As the former has stated: "If the child's parents are normally cold and unaffectionate, explicit evidence of this fact by punishment is nothing new to him" (1957, p. 334).

A recent dissertation by Sullivan, moreover, provided further empirical evidence for this supposition. The two major findings of his study in this context were (1) that mild verbal punishment served as an extremely effective reinforcement of behavior among kindergarten children and (2) that it was a far superior reinforcer than verbal regard alone or verbal reward and punishment combined (1960). Although his results related to concept elicitation, one could infer from the learning theory of socialization that his children, lacking supportive reinforcement histories, would not be conditioned to learn his simple task. On the basis of these findings in the literature and those by Sullivan, the writer developed the major hypothesis that susceptibility to mild verbal punishment would predict adjustment status among kindergarten children.

As in Sullivan's study (1960), the verbal punishment consisted of the phrase "that's bad" whenever a child selected a toy as his preferred response on a two-choice task involving a toy or a human.

PS 002102

The criterion for susceptibility to negative reinforcement was the increase in the number of human pictures selected following the punishment.

In brief, the theory was that the phrase "that's bad," would be more effective as negative reinforcement among adjusted kindergarten children than it would be among maladjusted ones. The full details in support of this theoretical position have been elaborated previously (Bommarito, 1964, pp. 17-24), and summarized by McCandless above. In any case, the desire of children to please adults has been well documented (Thompson, 1959). As already shown, learning theory explains this desire as a manifestation of healthy dependence on adults due to learned behavior. The experimental conditions of the study simulated, to a large extent, the socialization procedure adults frequently use to elicit desirable behavior in children since verbal punishment is such a prevalent part of that training. Consequently, in the absence of this desire to change responses in the simple discrimination learning task of the concept elicitation test under conditions of mild verbal punishment by an adult, the child was considered as not susceptible to influence from an adult and hence maladjusted.

To review, the major purpose of the previous study by Bommarito (1964) was to determine if susceptibility to mild verbal punishment by a significant adult figure would predict adjustment status. The verification of this hypothesis would serve both theoretical and applied purposes. On the theoretical side, it would provide construct validity for the extension of the learning theory of socialization to the clinical field. Moreover, the verification of the theory would embody a definition of emotional disturbance that is quite objective and testable. In short, the definition would be quite operational. As Bandura and Walters have repeatedly emphasized (1963), the clinical field is quite replete with value judgments on psychopathology and rather devoid of operational definitions.

With regard to applied aspects, the verification of the above hypothesis would serve as a first step in the development of a screening instrument for emotional disturbance encompassing the advantage of compactness, a theoretical orientation, and an operational definition. As indicated before, a compact screening instrument for emotional disturbance has yet to be developed and the screening process itself is crucial to the educational process of emotionally disturbed children.

In line with the learning theory of socialization, the previous study by the present investigator (1964) indicated that adjusted groups conditioned to the learning task to a statistically greater degree than did the maladjusted groups ($p < .05$). Before any firm confidence can be placed in the task as an effective screening instrument for emotional disturbance, however, a replication study for cross-validation purposes was sorely needed. Several writers have recently focused on the need for replications in psychological research. Milholland has even ventured to state that no one even does a replicate (1964, p. 320). Similarly, Matarazzo has declared that ". . . most clinical investigators continue to report successful conditioning of complex clinical phenomena." He goes on to say, however, that considerable progress would occur through replication (1965, p. 212).

The present study, then, was an attempt to cross-validate the previous findings by the investigator. If the cross-validation study verified the previous findings, further research efforts could not only refine the discriminative power of the instrument regarding adjustment status but also extend its range of application. An illustrated list of possibilities with respect to these areas could include the following:

1. From their study applied to schizophrenic patients, Salzinger and Portnoy demonstrated that patients who conditioned and (italics supplied) extinguished showed a very favorable prognosis for release from the hospital 180 days later (1964). Hence, an extinction procedure could be incorporated into the learning task.

2. The stimuli of the cards should be varied. At present, every card of the learning task has a picture of a toy on one side and a picture of a human on the other. These stimuli could be changed so that response differentiation under conditions of reinforcement could be made on the basis of two toys, for example, differing in the dimensions of shape, color, or use. If the learning theory of socialization is correct, the important issue is susceptibility to reinforcement, not differing reactions to specific types of stimuli.

3. The procedure of continuous reinforcement, as proposed in this study, may not be the most efficient one. As Hilgard (1956, p. 90) and others (Holland and Skinner, 1961, pp. 133-34) have intimated, intermittent reinforcement more closely simulates real life than continuous reinforcement. Hence, variations in schedules of reinforcement (fixed ratios, fixed interval, variable ratio, or variable interval) may yield more reliable data.

4. The results of the instrument could be correlated with Bower's techniques to determine if it does as adequate a screening with much less time and more convenience.

5. Variations in reinforcement contingencies may yield gain scores with significant age interactions since different age levels may have different reinforcement histories (Becker, 1962).

6. Some evidence suggests that extremely high gain scores could be associated with oversocialization as indicated by symptoms of

withdrawal, anxiety, and undesirable degrees of submissiveness (Bronfenbrenner, 1961). Finally, additional "offshots" could relate to the association between gain scores and numerous other variables, such as arbitrary power assertion by parents (Hoffman, 1960); the downward (Bettleheim, 1950) and upward mobility of parents (Swanson, 1961), and prognosis in therapy, casework, or special class placement.

In sum, the significance and value of the present screening instrument for learning theory in general and for the schools in particular should be clear. Accordingly, explanation may now be made regarding the general problems of the research and the specific variables possibly influencing the conditioning of the children's behavior.

II. SOME RELATED CRUCIAL ISSUES

In the writer's estimation, the most significant of the related methodological issues in this research were (1) the relationship between socioeconomic status (SES) and adjustment or mental health, (2) the measurement of SES, (3) the difficulties of ratings in general and teachers' ratings in particular, and, (4) perhaps most important of all, the operational definition of maladjustment these studies employed. The rationale for considering each of these variables as significant is described below:

The Relationship Between Socioeconomic Status and Adjustment

Several comprehensive research surveys have indicated that a more destructive parent-child relationship exists in lower class homes than in middle class ones; that controls are more likely to be harsh, arbitrary, punitive, and suppressive among lower class families; that a proper balance of warmth and support is much less likely among lower class families; that greater warmth, affection, and support exists in middle class homes; and that the middle class children generally show better adjustment and more favorable attitudes to authority. (Ausubel, 1958; pp. 351-52; Mussen & Conger, 1956, pp. 344-52; Sears et. al., 1957, p. 482; Mussen, 1960, pp. 442-43).

It is no wonder, then, that, on the basis of his summary of studies over a twenty-five year period, Bronfenbrenner concluded that the evidence conclusively shows "...that middle class parents, as compared with the lower class, are more permissive of the child's spontaneous desires, express affection more freely, and prefer 'psychological techniques,' such as reasoning or appeals to guilt, to more direct methods like physical punishment, scolding, or threats" (1961, p. 100).

In brief, the research on the relationship between socioeconomic status and adjustment appears to be rather definitive. It suggests very forcefully that children from the lower socioeconomic groups most likely constitute the majority of children with maladaptive behavior. Moreover, because of the emphasis sociologists place on this independent variable as a causal factor in behavior, some measurement of the relationship between social class membership and gain scores on the learning task was deemed necessary. Furthermore, Bower has reported that the results on his screening instruments are evidently independent of the social class factor per se but that certain occupational groups seemed to produce a higher incidence of emotionally disturbed children on his screening instruments. Families in which the fathers' occupations were classed as "service" and those classified as "semi-skilled," produced more than twice as many emotionally disturbed children as would be expected by their proportion of the state's population.

Families in occupational categories of "professional and managerial," "clerical and sales," and "unskilled," produced fewer than expected, (1960, p. 45). It should be clear, then, that Bower's findings have two implications for this study. First, since the social class membership per se did not influence the results on his screening instruments, gain scores on the simple discrimination learning task (the criterion of adjustment in this research) may also be independent of this variable. This again suggested the need to determine the social class of the child. Secondly, Bower's findings also indicated a relationship between occupational category of the father and adjustment of the child. Accordingly, some measurement of this relationship was made in this study.

Determination of Social Class

Numerous studies by Bower and others have shown that the father's occupation is a valid and reliable index for the determination of social class (Cook and Cook, 1960, pp. 93-95; Bendix and Lipset, 1953; Bergel, 1962). This index had the added advantage of facilitating a measurement of the relationship between occupational groups and incidence of maladjustment as previously outlined by Bower above. Consequently, each teacher was asked to complete a brief face sheet on each child she rated which required specific data on the father's occupation. In the original study (Bommarito, 1964), Bergel's classification of social class status (1962) based on paternal occupation was utilized for an approximate index of social class membership. The details of this classification are presented later. In the replication study, a refined quantitative measure of socioeconomic status was used. In this respect, detailed occupations were scored according to the combined average levels of education and income associated with a given occupation. The occupational classification system used was derived from the U. S. census of 1960 and consisted of 494 items, 297 of which were subgroups. The details and justification of this method have been described in full (U. S. Bureau of the Census, 1963). Previous work has indicated that the components of the occupational index--income and education--were highly stable over time, with test-retest correlation coefficients ranging respectively between .94 and .97 (Reiss et al, 1961, p. 152).

In any case, the paternal occupation was not classified in neither the original nor the replication study unless the data were complete and unequivocal. Only those cases with this type of data were included in the calculation of the statistics concerning the relationship between gain scores and social class status.

Difficulties in Ratings

Difficulties in ratings include drawbacks of ratings in general and those of teachers in particular. In the former case, Cronbach lists five possible sources of error. One is the generosity error, which is the tendency to give favorable reports. This probability will more likely occur if the rater believes the rating to be a reflection of his leadership ability. A second source of error stems from the ambiguity of the items or rating positions. A third common error is the biases of judges or constant errors, as reflected in the low inter-rater reliabilities. A fourth source of difficulty is due to the limited knowledge of the ratees by the rater. Finally, the halo effect may cause error in that ratings can be easily influenced by the overall impression of the ratee by the rater. In order to circumvent these possible sources of error, Cronbach makes a number of suggestions. First, the rater must be assured that the ratings are not a reflection of his ability. Secondly, the rater should know the ratee well, and a check should be made on this point. Thirdly, the scales should request, insofar as possible, ratings on specific descriptions of behavior, not on covert traits (1960, pp. 506-511). Whenever possible, then, these suggestions were incorporated in the teachers' ratings made for both investigations.

The agreement on general drawbacks of ratings is definite. Ratings by teachers presented a more controversial issue. Since these ratings served as the external criteria for the validation of the socialization theory and the screening instrument to be developed in this study, they need more careful attention. As White and Harris have pointed out and documented well, there are a number of disadvantages to teachers' ratings. First, teachers are prone to consider aggressive behavior as a problem. Secondly, inexperienced teachers and those near retirement age report more problems with their pupils. Thirdly, both the sex of the teacher and the sex of the child appear to have some effect on rater judgment. Fourthly, it appears that the teachers are more accurate in predicting delinquency than in predicting those who become mentally ill. Finally, considerable variability exists among teachers' judgments (1961, pp. 171-176).

Despite the possible drawbacks of teachers' ratings, there is good reason to believe that teachers are in the best position to rate pupils (Goertzen, 1957) - -a view which White and Harris have designated as not surprising when one considers the amount of working time they have together (1961, p. 173) - -that they agree very substantially with the highly acceptable criteria of diagnosis by psychologists, social workers, and/or psychiatrists who know the children intimately (Bower, 1958; Cooper et. al, 1959; Gildea, 1961), that they have very acceptable reliability coefficients (Beller and Turner, 1962, Peterson, 1961), and that they predict adjustment over time very well on a number of measures and quite possibly constitute the best single index for this purpose (Anderson, 1959).

In short, teachers' ratings of adjustment among children appear to be quite valid and reliable. Apparently, they also stand the test of time as well as any of our present diagnostic procedures with children. The question now remains as to the operational definition of adjustment. The application of such a definition as it applied to these studies may now be discussed.

An Operational Definition of Maladjustment

Before we can say that a child is emotionally disturbed, the criteria for this disturbance must be explicitly defined. Such a definition is crucial for the present studies. At the onset, however, one very important matter should be emphasized; namely, that these studies were not intended to establish an incidence of maladjustment. Considerable more research would have to occur with the hypotheses in these investigations before even beginnings in this direction could be made. Now that this precaution has been stated, an attempt to define the criteria of maladjustment of the research may be made.

On the basis of the foregoing presentation, it should be quite clear that there can hardly be any quarrel with the nature and meaning of learning task in the experimental situation under conditions of mild negative reinforcement. The theoretical foundations and empirical evidence supporting the major hypothesis from which this learning task springs are well-documented. The procedure for testing the theory is quite objective, communicable, and repeatable. And the evidence for learning is clearly denotative. In short, the concept of learning in the experimental task unquestionably meets the criteria for formal operational scientific meaning; whether it meets the second meaning of significance or usefulness for scientific concepts remains to be seen (Spiker and McCandless, 1954). This meaning, hopefully, occurred through concurrent validation with teachers' rating scales of established validity and reliability. The operational definition of these tools with their conceptual bases then provided the criteria for emotional disturbance used in these studies. It is this definition which now bears discussion.

In determining a definition of emotional disturbance a number of broad areas need consideration. Thus, the research supporting conformity and hence socialization as an index of adjustment is substantial:

1. Mussen and Conger cited considerable evidence from a variety of measurements which indicates that maladjusted children are not only less accepting of authority but also less likely to assimilate socially approved behavior so readily (1956, pp. 331-32).

2. Crandall's excellent study indicated that, among primary school age children, conforming children exhibited a number of admirable behavioral traits (1956).

3. Wright presented evidence which shows not only that patients "...could often withstand panic and malnutrition if they were in harmony with their community..." but further that the symptoms that recurred in such patients were also related to feelings of estrangement or isolation from their community. Community referred to family or peer group (1963, p. 290). If community harmony is so important to adults, it is logical to assume that it would be even more important to children who are so much more vulnerable.

4. Scott's fine critical review on research definitions of mental health and mental illness offers four lines of evidence in support of social conformity as a valid index of adjustment: (a) No other operational definition of mental health has any more evidence to support it; (b) the evidence validating the theory is substantial; (c) the measurement problems in the other approaches (e.g., psychiatric diagnosis, subjective unhappiness, and exposure to psychiatric treatment) present far many more difficulties; and (d) "...conformity to the social situation in which the individual is imbedded...would appear to be lurking in nearly all the other definitions..." (1958, p. 41).

5. Finally, as Hahn emphasized, the younger the individual the more society prescribes roles and the greater the force these regimented expectations of behavior have on the individual. As he further stressed: "One need not like...the enforced regimentations of the culture, but he must adapt (conform) for psychological, social, or even biological survival" (1962, p. 701).

In a related matter, the nature of the assessment techniques in these studies as these related to the proposed theory of socialization, bears scrutiny. "Ideally, in theory-oriented research, hypotheses under consideration should dictate behaviors to be studied and thus, indirectly, determine the choice of assessment techniques to be used" (Crandall, 1956). The theory presented here is that children with supportive and satisfying social reinforcement histories have learned to depend on adults and hence attempt to please them. Accordingly, they should be willing to please them, control their responses, and conform to the situation of the kindergarten. In short, they should be socialized and be susceptible to further socialization. Many items from the Problem Checklist (see Appendix A) and the Behavior Checklist (see Appendix B) primarily fit this theory. Since the teacher rating scales served as the external validating criteria for the adjustment designation of the child, several reasons existed in the cross-validation study for the use of these scales and an additional one, the Minnesota Scale.

1. Both the Problem Checklist and the Behavior Checklist contain subscales of withdrawn and aggressive behaviors, dimensions of behavior which are respectively regarded as manifestations of oversocialization and undersocialization (Peterson, 1961; Rubin et. al. 1966).

2. There is some evidence to suggest that the assessment of mental health may need a multidimensional approach (Scott, 1958). Both the Problem Checklist and the Behavior Checklist have subscales relating to aggressive, withdrawn, and immature behavior; whereas the Minnesota Scale (see Appendix C) uses a more global approach to the assessment of personality.

3. Because of the need for a multidimensional approach to the assessment of adjustment and the margin of error inherent in any measuring instruments of personality among children, the practice used by Bower to designate extreme groups in adjustment was further refined in these studies. By his system, a child was classified as maladjusted only if he scored at or below a certain rank on two of the three instruments. In the present studies, adjustment status was determined by extreme scores on all three scales or by a scoring system that was presumably the equivalent of this procedure.

4. As White and Harris have emphasized, four types of psychological problems are most likely to come to the attention of mental health personnel in the schools. These are "...the patterns of aggression,... social maladjustment,...educational maladjustment, and...behavior suggesting internal unhappiness (1961, p. 187). The scales used include items encompassing all of these behavioral indices. These, then, constituted the bases for the measuring instruments which served as the external validating criteria for the experimental learning task of these studies. These rating scales include behavioral items covering primarily the pattern of aggression and social maladjustment but also the manifestations of internal distress and educational difficulties.

In sum, the external validating criteria in this paper attempted to test the theory of this research, to validate it, to encompass a multidimensional approach, to cover patterns of behavior which are most likely to come to the attention of mental health personnel of the schools and which include most indices of mental health research studies in this field have utilized, and to maximize the accuracy of classification for emotional adjustment or the lack of it. Hopefully, two broad approaches achieved these purposes. First, each child was rated by his teacher by means of the Problem Checklist and two other instruments (the Minnesota Scale and the Behavior Checklist) with acceptable validity and reliability. Secondly, the child was rated as maladjusted or adjusted if he rated as such on extreme percentile points on all three scales. On the Problem Checklist and the Behavior Checklist, the items are slanted in the direction of undesirable behavior. Hence, the higher scores are indicative of maladjustment. The ten item Leikert-type Minnesota Scale, in contrast, requires that the teacher rate the child on a five point scale in the direction of adaptive or maladaptive behavior. On this procedure, the higher scores are in the direction of adjustment. In any case, the child was rated as maladjusted if he scored within the range indicating maladjustment on all three scales and as adjusted if he scored on the extreme percentile range indicating adjustment on all three scales. In order to simplify the statistics and to test the sex discriminative powers of the learning task fairly, these extreme percentile groupings applied to the two sexes separately. This procedure hopefully guaranteed an equal number of children from both sexes for the experimental task.

In brief, then, only extreme groups were included for the classification of adjusted or maladjusted. In line with most designations of extreme groups, these are represented respectively by the P_{27} and P_{73} cut-off points or their equivalents. Thus, on the Problem Checklist and the Behavior Checklist, children who scored at P_{73} or above were classified as maladjusted and those who scored at P_{27} or below as adjusted. On the Minnesota Scale, the reverse scoring procedure held true.

A formal operational definition of emotional disturbance may now be attempted. For the purpose of these studies, a child was classified as adjusted or maladjusted if he fell within the extreme percentile range indicating this dimension of behavior on all three teachers' rating scales which, in turn, sampled behavior in the areas of hostile aggressiveness, inability to adapt to a social situation, educational difficulties, and signs of internal distress as judged by teachers who knew the children intimately. The scales themselves already exhibited established satisfactory validity or reliability. In view of the wealth of evidence supporting the rationale and contents of these instruments, the comprehensive nature of their contents, and the number of possibilities they offer for the analysis of data, they appeared to have much merit. In the light of the apparent advantages of these scales, the details of which are discussed later, the above definition of emotional disturbance appeared to be quite satisfactory.

To recapitulate, it is hoped that this discussion clarifies an operational definition of maladjustment, describes the kinds of problems teachers are likely to refer to special service personnel; emphasizes the need for conformity in our society, particularly for very young children; documents some evidence on the validity of social conformity and adaptiveness as an index of adjustment, and indicates that the rating scales used contained items which not only fit these areas of definition, problems, and conformity but also encompass several indices of mental health or mental illness used in past research studies on these topics. In view of these factors and the fact that the direction of the results on these scales is clear for the designation of adjustment or maladjustment as defined previously, it follows, then, that sharply discriminating extreme scores on these instruments indicated considerable differences in adaptive or maladaptive behavior. It is these extreme groups which were tested by the experimental task.

One final note of caution deserves mention here to avoid a possible misinterpretation. The designation of a child as maladjusted by the criteria as used in these studies should in no way be interpreted to mean that, in the far distant future, the child will probably be referred for treatment, or be regarded as emotionally disturbed, or continue to behave or react as he does now. In other words, the predictive validity of the screening instrument which the writer attempted to develop must, at present, remain an open question. At this point, any conclusions on the stability of the above classifications or screening instrument should be quite premature. Further, the evidence on the stability of personality traits is quite controversial, contradictory, and confusing. For illustration, Thompson found little evidence for consistency in personality traits among children (1959, p. 20). Similarly, Becker failed to find much stability in ratings from infancy through adolescence, though the assessment of the love-hostility dimension among parents in the relationships with children held up well over time (1962). In contrast, the factor analysts have been quite optimistic regarding the stability of children's traits (Peterson, 1961). But their statistics leave much margin for error and their methods have been severely criticized for a number of shortcomings, including an inadequate appreciation of momentary environmental factors, a lack of any contribution of new ideas, an atheoretical orientation, and unselectivity in the choice of material to be subjected to factor analysis in the first place (Atkinson, 1960, p. 278). Moreover, in what is probably the most comprehensive study on the prediction of adjustment over time, predictions for adjustment fared much better than those for maladjustment (Anderson, 1959). Furthermore, the difficulties in predicting consistency in behavior over time are, it seems, to be expected for several reasons. First, "...since the child's personality is relatively unformed and is changing rapidly, there is obviously less basis for consistency than in the adult" (Ausubel, 1958, p. 101). Secondly, learning theory "...tends to...regard the human being as modifiable at any age." It probably has a greater tendency in this direction than any other theory (McCandless, 1961, p. 322). Some evidence would support this assumption. Examples are the studies by Lang (1962) and Bandura (undated) indicating that changes in behavior can occur by proper social reinforcement without any necessity to determine the "cause" of behavior, which may be quite remote and exceedingly difficult to determine, and the documentation by Ellis (1959) that, in all probability, no particular condition or set of conditions is absolutely necessary for basic personality change in that numerous situational and life factors can produce such change. Finally, the problems in determining consistency are enormous and multiple and the possible causes for stability or change in personality are numerous. Ausubel has discussed these in detail (1958, pp. 100-103, 264-67). The point of all this discussion has been ably summarized by White and Harris:

"The criteria for detecting psychological difficulties are pragmatic ones based on the definition of maladjustment as currently abnormal behavior--compared to that of the current group in the current situation. We are not implying that there is necessarily a relationship between the difficulties so detected and later emotional disorder. Whether such a relationship exists or not, or to what degree, is a crucial question...to answer from more adequate data than we now have" (1961, p. 187).

It is hoped that the research discussed above clarifies and justifies this point of view. It is also hoped that the present research studies offer some of these data in question. In any event, it should be clear that the comments by White and Harris on the criteria for psychological difficulties represent the viewpoint that this writer adopted for these investigations.

III. HYPOTHESES AND SAMPLING PROCEDURE

In the original research (Bommarito, 1964), as in the present cross-validation study, the major operational question was whether or not the gain scores on the experimental task under mild verbal punishment would discriminate the groups designated as maladjusted from children classified as adjusted. In the original investigation, the adjusted groups showed greater gain scores at a statistically significant level ($p < .05$). A number of subsidiary but related operational questions were also tested.

1. Would the block four gain scores for the group as a whole show a statistically significant increase in the desired criterion of human preferences when compared to the basal score at the end of block one?

2. Would the correlation of block four gain scores to mental age on the Peabody Picture Vocabulary Test depart significantly from a zero order correlation coefficient (Pearson r)?

3. Would the mean gain scores from block one through block four under different examiners show statistically significant differences?

4. Would significant interaction effects between examiner and the adjustment of the child occur with respect to differences in gain scores?

5. Would the gain scores on block four show statistically significant sex differences?

6. Would paired comparisons between groups of children of different social classes, as determined by paternal occupation, indicate statistically significant differences in gain scores?

7. Finally, would gain scores of children be independent of the type of paternal occupational category?

These, then comprised the major operational questions of the investigations. The findings of the original research concerning these respective questions may now be summarized (Bommarito, 1964, pp. 66-84).

1. A $\underline{t}$ test for uncorrelated samples applied to the entire sample (N of 60) to determine the significance of differences in the mean number of human preferences (raw scores) prior to conditioning and subsequent to this process yielded a value of 12.29 ($p < .0006$).

2. The Pearson-Product correlation between mental age and gain score, with chronological age controlled, was negligible (r of .07).

3. When the gain scores for the entire sample were studied, no significant examiner differences emerged (t of 1.46 with p almost .16).

Significant differences did occur, however, between the gain scores of maladjusted children under the two examiners (t of 4.14 with p less than .001). Moreover, under the two female psychologists, the difference in mean gain scores between the examiners seemed to be largest with the male-maladjusted group, particularly in the third and fourth blocks.

4. A two-way analysis of variance revealed that the joint effect of the examiner and the adjustment status of the child (interaction effect) was decidedly insignificant (F value of .60 with $p > .75$).

5. No significant differences in average gain scores occurred between the two sex groups (t of .89 with $p > .32$).

6. Using Bergel's four-fold classification system (1961, pp. 270-72) of upper middle class (Class I), white collar group (Class II), upper lower class (Class III), and lower social class status (Class IV), no significant differences in mean gain scores on Block IV existed between any two social classes (Mann-Whitney U Test). In support of Bower's finding, however, the mean gain scores for Block IV by children with fathers in the occupational categories of professional and managerial, clerical and sales, and unskilled far exceeded these scores for students with fathers in the occupational categories of service and semi-skilled. Thus, by testing an a priori hypothesis that no significant departure from a 50-50 ratio of scores within each group in excess of a critical cut-off scores would be found, a chi-square value of 6.074 was obtained ($p < .02$). Two-tailed hypothesis testing was used in each hypothesis above since insufficient prior data existed to predict the direction of the difference in advance.

Accordingly, these were the major operational questions to be tested in the present replication research. In addition to these major operational questions, other matters of significance in the research design were the nature of the universe and the sampling procedure of the two studies.

Universe. In the original study (Bommarito, 1964), the universe consisted of all public school children who had entered kindergarten during September, 1963, in a large all-white suburban area of Detroit. The school-age population consisted of approximately 1500 children which were quite evenly divided by sex. A random sampling of these children were rated by 27 teachers. The community itself was a rather prosperous and homogeneous all-white population with nearly 50,000 inhabitants.

In Bergel's social class scheme (1961), the adult inhabitants could be primarily classified as white collar and blue collar groups (Bommarito, 1964).

In the present cross-validation research, the universe comprised all children who had entered kindergarten in September, 1967, in a heterogeneous community of about 37,000 inhabitants, including such elements as two universities, a large industrial complex, and mixed racial groups. Five of the schools received federal aid because of a large number of families who qualified as culturally deprived (incomes of less than $3,000.00 per annum). The school-age population included 721 children with a preponderance of males (53.5 percent boys vs. 47.5 percent girls). A random sample of the population was rated by 16 teachers. Under Bergel's classification scheme, the social classes were much more heterogeneous than in the previous study, with a greater incidence of professional workers.

Sample and sampling procedure. The details of the previous sampling procedure (Bommarito, 1964) were essentially replicated in the present study with a few minor revisions because of the large difference in the size of the populations. The essential elements have been summarized below.

1. A complete roster of the school system was obtained.

2. Each child was assigned a random number.

3. A table of random numbers was entered to select the children who would be evaluated by the teacher rating scales. In selecting the children, however, strata were provided for sex groups and teachers so that an equal number of boys and girls was chosen and each teacher had the same load (seven children) from each sex classification.

4. A sufficient number of alternates were also selected randomly (14 percent) to provide for complications from sample loss due to transfers, absenteeism, illness, etc., without seriously biasing the sample. Further, this plan allowed for a sufficient number of children in each sex stratum (n of 112) to provide adequately large experimental groups for the learning task on the basis of the upper and lower twenty-seventh percentile points which served as the cut-off scores for the designation of extremes in adjustment status. Thus, each group had 30 children classified as adjusted and an equal number designated as maladjusted, while the total sample of original ratees and alternates included 60 children in each adjustment category.

5. The foregoing procedure provided an N of 30 for most of the crucial variables of the research. As Smith has pointed out, an N of this size has some distinct advantages. As he stated, "...the critical ratio...tends to be distributed in a pattern which follows the normal probability curve provided the combined number of cases in the two samples is not less than about 30" (1962, p. 74). According to Smith, then, the selection of 30 cases would support hypothesis testing for differences between a number of significant variables: on gain scores between sex strata ($n_1 = 30$, $n_2 = 30$), differences between the adjusted and maladjusted groups ($n_1 = 30$, $n_2 = 30$), intra-sex differences

($n_1 = 15$, $n_2 = 15$), examiner differences ($n_1 = 30$, $n_2 = 30$), etc. In short, an N of 30 for each of these important measurement areas provided for more effective hypothesis testing of significant factors which could have influenced the gain scores by the use of the critical ratio t in a pattern following the normal probability curve. In other words, the t ratio scores were, in essence, z scores and could be interpreted in this manner.

It should be reemphasized, however, that for the learning task individually administered by the two psychologists the plan was for a 100 percent replacement for each of the four adjustment classifications which had identical numbers of 15 original examinees. The alternates thus provided for replacement losses well above the level Sullivan (1960, p. 12) found necessary for his large metropolitan area (23 percent replacement) where the likelihood of sample reduction was conceivably greater than the two communities under consideration here. The results, in fact, indicated not only the validity of the foregoing point but also that very little turnover occurred for both the teachers' ratings and the individual examinations by the two psychologists. The table below embodies the data of both studies:

TABLE I

Number of Alternates By Sex and Conditions of Testing

	Teachers' Ratings	Individual Testing	Totals
Males	9 (3)	1 (2)	10 (5)
Females	5 (8)	2 (1)	7 (9)
Totals	14 (11)	3 (3)	17 (14)

Since there were 200 ratees for the teachers' scales and 60 individual examinations, the turnover rates in the original study were respectively seven and five percent. The statistics in parentheses refer to the turnover rates in the replication study which had 224 original ratees for the teachers' scales and 60 original examinees for the individual examinations. The replacement rate in each instance was five percent, even less than the original investigation, a somewhat surprising finding in that the former was executed in a relatively prosperous suburban area as compared to a racially mixed city with culturally deprived sections.

As a final note, prior to the individual examinations, the 15 children and the 15 alternates in each of the four classifications were randomly assigned to the two examiners. The end result of this procedure was that each examiner had exactly the same number of children to test in the broad adjustment categories (adjusted and maladjusted) and approximately the same number in the four classifications mentioned above (male-adjusted, female-adjusted, male-maladjusted, and female-maladjusted).

IV. METHOD AND INSTRUMENTATION

This section includes conditions for testing, as well as the methods and instrumentation which were used in the investigations.

Conditions for Testing

The first task of the research worker was to outline the study in its broad and general form to appropriate top school authorities, obtain their permission to conduct the study, and arrange an open meeting with all the elementary principals and the kindergarten teachers (see Appendix F for step by step progress of research). At the open meeting, the roles of the teacher and the nature, time, and conditions were explained (after Sullivan, 1960, p. 40). Participation in the study was sought strictly on a voluntary basis. As in the original study, all teachers agreed to participate. After this agreement, each teacher received a complete packet (see Appendix F), including the appropriate number (14) of the first rating scale, the Problem Checklist. With each rating scale, teachers were further requested to complete a brief fact sheet on each ratee. With the first rating scale, teachers were also requested to complete a brief questionnaire on background data (see Appendix G). The tests used and their order of use were the Problem Checklist, the Minnesota Scale, and the Behavior Checklist. Following the statistical analysis designating the children as adjusted or maladjusted on the basis of extreme scores on all of the three scales, the subjects in each classification were randomly assigned to the two psychologists for individual examinations. Finally, prior to any ratings, it was stressed to the teachers that the children had been selected randomly, and that the choice of any particular ratee was strictly a matter of chance. These tasks, then, related primarily to the teachers' responsibilities.

The individual examinations were done by two well-trained male psychologists. One had received his M.A. in school psychology, while the other individual had just obtained his doctorate in clincial psychology. At any rate, before the individual examinations occurred, a short notice was sent to the principals and the teachers as a reminder that these would occur and had administrative support (see Appendix H). At approximately the same time, a similar note was sent to the parents of the children who were to be examined individually(see Appendix I). A complete testing schedule accompanied these notices forwarded to school personnel. In conducting these individual examinations, moreover, an attempt was made to control for the important variables in research work with children relating to the conditions surrounding the testing and to the testing procedure itself (Bijou and Baer, 1960, pp. 174-75).

1. Within each school, the examiner personally came to select the ratee. The name of the child and the teacher was on a slip of paper and

given to the latter. In selecting the child, the examiner had been instructed to explain the task in a matter of fact but pleasant way (i.e., "it is your turn to play games"), to be neutral in explanations of the task, and to answer the child's questions pleasantly but not to initiate conversation. He was also cautioned to avoid involvement with the teacher.

2. Several points were particularly emphasized to the teachers and administrators. First, advanced preparation of the children was neither necessary nor desirable. The teacher was simply requested to say that Mr. __________ would be spending some time with the children in the room. Secondly, the teachers were urged to assure the children that playing the games with Mr. ________________ had their approval. Thirdly, teachers were discouraged from doing any questioning of the child after he returned to the room on the grounds that this could conceivably contaminate the results. Finally, the writer promised that the results would be made available to the school personnel. This will be done as soon as possible.

Method

After an introduction of himself and a brief query as to how the child would like to play games, the examiner asked the child a few questions on his name, age, grade, and birthdate. A few minutes were spent in establishing rapport with the child. Toys were used to facilitate this process. As Sullivan stressed, the examiner had to proceed slowly (1960, p. 13). Once the child appeared to be as comfortable as possible with the brief period allowable, the Peabody Picture Vocabulary Test was administered. The learning task then followed these psychometric measurements. "The concept elicitation test (or the learning task) consisted of twenty-five cards measuring 4" X 14". Two drawings in black India ink, one of a toy and other of a person, appeared on each card. The toys varied from a simple ball to a slide, and the persons varied in age, sex, and activity depicted. The drawings were randomly assigned to the left or right of the cards" (Sullivan, 1960, p. 59).

The method of administration for this task was as follows: The examiner placed the stack of experimental cards before the subject and read instructions to the child as Sullivan has outlined them:

"Now we are going to play another game. I have some cards with pictures on them. I am going to show them to you one at a time and all that I want you to do is point to the one you like the best (E said, 'Point to the one that you like best' periodically). Before the second block of trials, E said, 'Remember, just point to the one that you like the best'..." (Sullivan, 1960, p. 13).

As in Sullivan's study, the child was given four blocks of 25 trials each. On the first block, the child was simply requested to point to the picture he liked best. The number of times human drawings were selected on this first trial constituted his base or basal score. Beginning with the second block of trials and continuing through the fourth block, the child's responses were reinforced by mild verbal punishment in which the E would say, "that's bad" each time the S pointed to a toy. To eliminate initial variability on the base score between the sexes and among young children, Sullivan's plan was emulated. Under this plan, "...the Block I score was subtracted from the succeeding block scores; a constant of 25 was then added" (Sullivan, 1960, pp. 13-14). Finally, Sullivan's study also indicated that "...reinforcement ... produced cumulative effects over blocks." It followed from this assumption that the most satisfactory measure of the conditions effect would be the means on Block 4, where these cumulative effects would be maximized (Sullivan, 1960, pp. 17-18).

In sum, the Block I score for any child was subtracted from his Block 4 score and a constant of 25 then added. The resulting sum was his gain score for Block 4 or his cumulative gain score over blocks. It was on the basis of this measurement scheme that the "gain scores" on Block 4 and their means were calculated. In addition to this measurement process, other important factors with regard to the procedures used with the children by the psychologists were as follows: First, no attempt was made to delude the child as to the purpose of the experiment; the child was presented with an opportunity to ask questions before he returned to the classroom. Secondly, the examiners did not know the adjustment classification of the children. Finally, since two previous studies had shown that children were not aware of the reinforcement contingencies for their responses (Sullivan, 1960; Bommarito, 1964), no check on this awareness was made in this replication study.

Instrumentation

The materials used for this study included the following:

1. A brief face sheet on each S to be rated requesting data on his name, sex, birthdate, father's name, father's occupation, and the length of time the teacher had known the child.

2. Three teacher rating scales, including the Problem Checklist, the Minnesota Scale, and the Behavior Checklist. The statistical support for the first scale, the Problem Checklist, rested on interrater reliability, internal consistency, and low intercorrelations between subscales. In the first case, dual ratings by kindergarten teachers yielded the respective reliability coefficients of .77 and .75 for the subscales of conduct disorders and personality problems. In the second instance, the high factor loadings for the subscales suggested internal consistency. The correlations between factors, moreover, was .18, low enough to meet most requirements for independence. Finally, the nature of the two principal factors of the scale can be clarified by definition. In this

connection, the factor labeled "conduct problem" implied a tendency to express impulses against society. The second factor, called "personality problem" contained a variety of elements suggesting low self-esteem, social withdrawal, and "dysphoric" mood (Peterson, 1961).

The second rating scale used, the Minnesota Scale, has been found to hold up well in predicting adjustment in several areas for both sexes over a seven year period with highly statistically significant correlations to external validating criteria. For illustration, the results of this rating scale were found to correlate .52 with records of participation in school and community, .44 with ratings of prominent citizens who had known the subjects in various roles, and .52 with a score combining several of these rating devices. These correlations were highly significant statistically with a probability of chance variations in the sample as the cause of the results to be less than once in a hundred times (Anderson, 1960, pp. 55-57). It is obvious, then, why Anderson thought so highly of this instrument: "The teacher's rating blank ... seems to be an excellent instrument in terms of long-term prediction ....in its ten-item form ... it holds promise for development as a screening device." Among the many advantages this scale offered, Anderson mentioned "... the ease of administration, the short time needed for its use, and its predictive value ..." The scale has been used with young children at early elementary (Anderson, 1960, p. 67) and kindergarten levels. Though the results of the latter have not been completely analyzed, preliminary analysis has indicated that they appear to be promising (Gram, 1962).

The validity of the last teacher rating scale, the Behavior Checklist, rested largely on internal consistency based on factor analytic studies. The variables of the checklist can be subsumed under seven major factors. In order of decreasing importance, these are (1) disorientation and maladaptation to the environment; (2) anti-social behavior; (3) unassertive, overconforming behavior; (4) neglect; (5) infantile behavior; (6) immature social behavior; and (7) irresponsible behavior. The meanings of these factors have been described in detail by the authors (Rubin et. al., 1966). Though no data on the test-retest reliability coefficient of the scale previously existed, the present author obtained this information for 53 children in which the same teacher made dual ratings on the children in her class over a three week interval. The coefficients obtained were .94 for the entire sample and .80 for the maladjusted ($p < .01$ in each case). Finally, at the suggestion of Rubin himself, in the interest of improving the reliability of the scale, the original scoring system of the scale which simply required the teacher to check items of undesirable behavior applicable for a given ratee was changed. As Rubin suggested, the items might have a scale of several parts associated with them, "thus requiring the rater to focus attention on each and every item rather than its current form, checking those items only that appear to be significant" (Rubin, 1966). This change was incorporated in the scoring system which required the teacher to rate the severity of behavior on a scale ranging from 0 (no problem) to

2 (a severe problem). This scoring system had the additional advantage of equating it with the one employed on the Problem Checklist. Finally, the ratings on the BCL in this investigation were found to be independent of (1) levels of education (2) years of teaching experience, and (3) a combination of education and experience as measured by the Mann-Whitney U Test (Popham, 1967).

These data, then, provided the validating support for the teacher rating scales used in this investigation as the external criteria for designating the children adjusted or maladjusted from P_{27} or P_{73} cut-off points (or their equivalents) on all three scales. In the original study (Bommarito, 1964), the adjustment status of each S was determined on the basis of extreme percentile cut-off points from the raw scores. In the present investigation, however, this scoring procedure did not provide the desirable replacement rate (100 per cent) for the children to be examined individually on the learning task. For this reason, normalized standard scores, as described by Popham (1967, pp. 36-37), were used as the scoring method. "Such . . . scores can be used to average students' scores on different tests since they have a common base" (Popham, 1967, p. 34). The scoring procedure, then, was to classify children as adjusted if they averaged P_{23} or less on all three scales or as maladjusted if they averaged P_{73} or more on all three scales. Within the maladjusted group, the Ss were assigned in order of high score to low. Within the adjusted group, the Ss were designated in order of low score to high. The method had the advantage of providing 100 per cent replacement rates for each adjustment classification (female-adjusted, male-adjusted, female-maladjusted, and male-maladjusted) but it also had the dubious feature of eliminating the assignment of some children to the learning task who had obtained extreme cut-off scores on all three scales. Such was the scoring procedure used in this replication study.

In addition to the brief face sheets and the rating scales, the instrumentation and procedures also included the following:

1. The occupational classification system derived from the U.S. Census of 1960 previously described.

2. Bergel's social status classification system. According to this author (1962, pp. 270-72), the hierarchy of occupation, from those with the highest prestige to those with the lowest prestige, is as follows:

Class I: This group includes managers, professionals, and proprietors. These constitute the upper middle class.

Class II: Clerical and sales positions occupy the lower middle class. Both Class I and Class II may be designed as the white collar group.

Class III: The bulk of this classification is composed of skilled workers and craftsmen. These may be designated as belonging to the upper lower class, and as the blue-collar group.

Class IV: Service jobs, semi-skilled and unskilled positions comprise this lower social status classification.

3. Two Concept Elicitation Tests provided by Sullivan which served as the conditioning tasks of the study.

4. The Peabody Picture Vocabulary Test.

V. RESULTS

As a parallel exposition to the description of the methods, the results will be presented in relationship to (1) the entire sample of 60 children, (2) adjustment classifications, (3) sex and examiners, (4) intelligence, (5) social class status, and (6) some clinical findings.

Findings on the Entire Sample

Perhaps the first step in the analysis of the data was to determine if conditioning by mild verbal punishment (negative reinforcement) had any effects on the responses of the children. If the results proved to be negligible, the value of the entire study could be questioned. For this purpose a *t* test for uncorrelated groups was applied to the entire sample to determine the significance of differences in the number of human preferences prior to conditioning (beginning of Block I) and subsequent to this process (end of Block 4). The *t* value of 12.22 for changes in these raw scores was statistically significant at a very high level ($p > .0005$ in a one-tailed test). Since a very large *t* value for a similar measurement (t of 12.29) had been found in the original study (Bommarito, 1964), the one-tailed test was justified.

Subsequent to this *t* test, a one-way analysis of variance was applied for the weighted mean gain scores over the three blocks of reinforcement. This procedure yielded some distinct types of data. In this respect, it disclosed significant variations between blocks and within subjects (see Table II). It also made it possible to determine the maximization of block effects. As indicated before, both Sullivan (1960) and Bommarito (1964) found that the maximum conditioning effect was over three blocks of reinforcement. The present study also tested this reinforcement effect. In the analysis of variance itself, the procedure for correlated groups was used (Garrett, 1962) in that the statistic applied to the same sample under identical treatment. The results indicated highly significant differences in the mean weighted gain scores between blocks and among subjects ($p < .0005$ in each case for one-tailed tests of the *F* ratios). Again, similar results in the original study (see Table II) justified the use of a one-tailed test.

In addition to this principal finding of significant treatment effects and individual differences, Garrett has suggested that through a further analysis of variance it can be determined whether individual differences in gain scores (F ratio for subjects) over the three blocks of reinforcement are significantly greater than differences due to the amount of negative reinforcements (Treatments). Thus, if *V subjects* are divided by V *trials*, the resulting F is 113.67/310.72 or .3757. For df of 59 and 2, a F table reveals that the .05 level of significance is 19.47 or more. Hence, in the present study, it can be concluded with considerable confidence that the individual differences in gain scores were no greater than differences due to the amount of negative reinforcement. As Garrett has pointed out (1962, p. 395), the reverse is usually true. Accordingly, the implication of the data was that the effect of negative reinforcement between blocks on gain scores must have been quite drastic, a conclusion substantiated by the extremely large F ratio for treatment, by an F ratio between trials equally as large as that among subjects, and by the results of the t test between uncorrelated samples where the unweighted scores for human preferences before conditioning were compared to the same type of scores after three blocks of negative reinforcement.

The question of the maximization of block effects has, however, not been solved. In this connection, Appendix K should be studied for a complete discussion on this problem. In general, a test for significant differences in variances found the groups to be highly homogeneous. With this homogeneity present, it was possible to apply the within groups variance (see Table II) as an estimation of population variance to all possible pairs of means for gain scores between blocks. The following table summarizes the results:

TABLE III

Significant differences in weighted mean gain scores between blocks of reinforcement for all possible paired comparisons in the 1968 sample

Block Gain Scores Compared	Means Compared	Difference		Levels of Significance
Bl. IV vs. II.	36.383-31.850	4.533	Much	$< .0005$
Bl. IV vs. III.	36.383-34.467	1.916		$< .0005$
Bl. III vs. II.	34.467-31.850	2.617		$< .0005$

It follows, then, that the present study confirmed the previous findings by Sullivan (1960) and Bommarito (1964) on the maximization of effects from negative reinforcement; namely, that the peak of these effects occurs after three blocks of reinforcement (see Table IV).

In summary, findings on the entire sample disclosed (1) that the effects of negative reinforcement were quite powerful; (2) that the changes in gain scores were as much attributable to differences in the amount of reinforcement as they were to differences among subjects; and (3) that maximization of the influence on gain scores by negative reinforcement occurred after three blocks.

Classification of Adjustment

The principal hypothesis of this study was that the mean weighted gain scores on Block 4 would discriminate the children classified as maladjusted from those categorized as adjusted. But the possibility existed that these final differences may have been due to initial differences in the number of human preferences made prior to any conditioning in the two groups. Thus, if the adjusted group had a significantly greater mean number of human preferences on Block 1 a subsequent superiority on the mean weighted gain scores of Block 4 would be difficult to interpret. But a *t* test for significant differences between the two groups on the mean raw numbers of human preferences made on Block 1 revealed no significant differences for the t of .53 could have arisen from chance variations more than 70 times in a hundred (Dixon and Massey, 1957, p. 384). Patently, the number of humans preferred prior to conditioning had little influence on any differences in mean weighted gain scores between the two adjustment groups. These results closely paralled the findings in the original study (Bommarito, 1964, p. 69). With the influence of this variable eliminated, the examination of the data relating to the major hypotheses was facilitated.

In the original study (Bommarito, 1964, p. 69), a *t* test for significant differences in mean weighted gain scores for Block 4 revealed a t of 1.961 with a probability of chance findings less than .05 ($p < .05$). Further, the differences in means was substantial (36.20 for the adjusted vs. 32.01 for the maladjusted). The conclusion was, therefore, that, under conditions of verbal negative reinforcement, the adjusted group showed a greater performance on the criterion response (number of human pictures selected on Block 4) than the maladjusted group at a statistically significant level. Because of the prior verification of this hypothesis, a one-tailed test was applied in the present replication study. The *t* value obtained for this statistic was 1.04. In a one-tailed test with 58 df, such a value has a chance probability more than .20 ($p > .20$). Consequently, the conclusion was to accept the null hypothesis of no significant differences between the two adjustment groups on the mean weighted gain scores for Block 4. The lack of statistically significant differences between adjustment groups was also emphasized by the small F ratio for rows in the two-way analysis of variance (see Table V).

At this point, the method of assigning subjects to adjustment groups should be reemphasized. It will be recalled that the children were assigned to either adjustment groups on the basis of *average* rankings derived from normalized standard scores. The statistical justification for this procedure has been explained and the rationale was to guarantee a sufficiently large number of replacements in the event of a large "turnover" for original ratees. In the original study (Bommarito, 1964), children were assigned to adjustment groups on the basis of extreme cut-off scores on all three scales. In the investigator's opinion, then, the results of the present study were possibly misleading because of this procedural difference. For this reason, a statistical check was made on this possibility.

In terms of the foregoing reexamination, the data were analyzed by comparing the gain scores on Block IV for all children in the adjusted or maladjusted categories who respectively scored at the P_{27} level or less or at the P_{73} cut-off point or more on all three scales. The calculation resulted in an N of 24 for the maladjusted group and the retention of the original sample for the adjustment classification (N of 30). Subsequently, the original mean weighted gain score on Block 4 for the entire group (36.38) was used as the critical cut-off point to determine the relationship between adjustment category and the number of scores above or below the mean. If no relationship existed between membership in these two groups and the number of scores below or above this critical division score, the expectation should be that half the scores in each group would exceed the cut-off point and half would not. In brief, a test was made on an a priori hypothesis that the ratio of scores within each group exceeding or not exceeding the original common mean would be 50:50 or that no statistically significant departure from this ratio would be found. With correction for continuity (Popham, 1967), the X^2 value of 4.20 with one df, however, was highly significant (with p between .05 and .02). This finding thus provided partial replication for the original finding that adjusted children would show a significantly greater mean weighted gain score on Block 4 under mild verbal punishment than the maladjusted group would. It also suggested that the failure to find a significant difference on the mean weighted gain score between these two adjustment groups rested, in part at least, on procedural error.

Since the replication study provided at least partial support to the major hypothesis, one final matter merited consideration. Reference here is to the relationship betwen the subscales of neuroticism or Personality Problems and Conduct Problem (see Appendix A for Subscales). The tentative hypothesis was that the former possibly reflects oversocialization, the latter undersocialization. If the gain scores were sensitive in detecting socialization skills, children in the former category would show higher gain scores than those in the latter groups. Conversely, children with greater scores in conduct problems than in neuroticism should show smaller gain scores on Block 4 than students primarily classified in the latter nosology. As a partial check on these relationships, two statistical calculations were made within the maladjusted group. First, the weighted gain scores on Block 4 for children who scored above the mean on the Neurotic Scale but below the mean on the Psychopathic Scale (classified as neurotic) were compared with the same gain scores of children who scored above the mean on the Psychopathic Scale but below the mean on the Neurotic Scale (classified as Conduct Disorder).

Under the Mann-Whitney Test, the U value of 53 found far surpassed the U of 33 or less needed in a two-tailed test for the .05 level of significance. The null hypothesis of no significant differences in gain scores between the two groups was therefore accepted.

The problem on the relationship between gain scores and performance on the Problem Check List (PCL) by maladjusted children was also analyzed from a different perspective. In this sense, the subscale scores of the PCL were compared for children who scored one S.D. or more below the mean on the gain scores or one S.D. or more above the mean on the same data. In the former case, the Wilcoxon Test for matched samples yielded a T score of 24.5 (N of 11), far beyond the value of 11 or less required for the .05 level of significance. Similarly, for children scoring well above the mean on the learning task (N of 10), a T value of 22 was obtained, which was far in excess of the score of 8 or less required for the .05 level of significance on the Wilcoxon Test. In short, within the limitations of this study, no relationship was found between extremes of performance on the learning task and patterns of scores on the PCL.

By way of momentary digression, this last mentioned conclusion should be discussed immediately to avoid the elimination of a potentially useful hypothesis, namely, the relationship between type of personality adjustment and the nature of socialization which has developed, suggesting that neurotic children are overinhibited and that conduct disorders lack controls. To this end, several preliminary "leads" can be offered for failure to find a relationship between the measure of socialization in this study (weighted gain scores on Block 4) and the classification of adjustment pattern (the Subscale scores of the PCL). The list is only suggestive, of course, and possibly can be elaborated. Nevertheless, the reasons given below for failure to find this relationship possibly serve as an excellent starting point for the investigation of the problem under consideration:

1. The theory itself is questionable. The large body of research supporting this formulation makes this objection doubtful (Quay and Quay, 1965; Quay, Morse, and Cutler, 1966; Quay, Sprague, Shulman, and Miller, 1966).

2. Neither instrument is valid for what it purports to measure. Since evidence exists for both instruments as indicators of maladjusted behavior, this objection would be premature at this point but still should be carefully pursued. Despite this precaution, a number of relevant observations may be made. First, over three separate studies (Sullivan, 1960, Bommarito, 1964, and the present investigation), no test-retest reliability coefficients on the learning task exist. Secondly, though test-retest reliabilities were .77 for the conduct disorder and .75 for the personality problem dimension on the PCL, these statistics refer to dual ratings made by the same kindergarten teacher on her children. When different kindergarten teachers provided the same

ratings, the test-retest coefficients dropped to .52 for conduct problem and .38 for personality problem (Quay and Peterson, unpublished, 1967). Thirdly, the atheoretical nature of the factor analytic method should be reemphasized at this point. Scales are derived from the intercorrelation of items, not necessarily for the theoretical or logical nature of what the items are supposed to measure. Thus, it would be quite possible for teachers to have high test-retest coefficients in an entire scale or on subscales without deriving scores which sharply distinguish a child as to personality type. This phenomenon even occurred in the present investigation. Thus, the means for neurotic disorders and the Conduct Problems on the PCL were respectively 7 and 6, based on the scoring system suggested by Quay and Peterson (unpublished, 1967) in which the score is simply the total number of unfavorable items checked.

3. Closely related to the matter of validity is the problem of sampling. In this respect, the number of items in each subscale is rather limited, with 17 items included in the Conduct Problem Dimension and only 14 represented by the Personality Problem Factor. Similarly, the samples in the present study on which the relationship between subscore patterns of the PCL and scores on the learning task was checked were quite small, ranging between 10 - 11 subjects.

In short, the primary problems which make it difficult to evaluate the failure in finding a relationship between performance on the learning task and the subscale patterns on the PCL relate primarily to questions of validity, reliability, and sampling of behavior. Finally, an additional potent observation may be made that the very young child has not developed habitual types of behavior to the extent that one or another set of coping mechanisms predominate in a generalized fashion. As behaviorists have emphasized, moreover, (e.g., Bandura and Walters, 1963; McCandless, 1961), situational events have a profound influence on behavior for behavior changes as reinforcement contingencies vary. Even cognitive theorists believe that the child's personality is in a constant state of flux (Ansubel, 1958).

To summarize this section on the relationship between gain scores and adjustment classification, the replication study did not fully verify the original finding of a sharp differentiation in mean weighted gain scores on Block 4 in favor of children in the adjusted group. When statistical adjustments were made on the assignment of the children to adjustment categories, a chi-square test did favor the gain scores for the adjustment group at a statistically significant level. The discrepancy in findings suggested a procedural error and at least partially cross-validated the originally verified hypothesis. Finally, no relationship was obtained between the patterns of the subscales on the PCL and gain scores on the learning task. Possible exploratory explanations offered for this failure included several statistical shortcomings regarding the instruments and the adequacy of the sampling, as well as the age of the child.

Sex and Examiner Differences

With one exception, the sex of the child and the specific examiner who administered the learning task had little influence on the gain scores. Specifically, the findings were as follows with respect to each variable under sex or examiner differences in gain scores: (a) sex of child (t of .16, df 28, $p > .20$); (b) examiner differences for the entire sample (t of 1.39770, df 28, p between .20 and .10); (c) examiner differences for maladjusted children (t of .85, df 13, $p > .20$ in a one-tailed test); (d) examiner differences for adjusted children (t of 1.14, df. 28, $p > .20$); (e) interaction data (see Table V), testing the joint influence of the sex and adjustment status of the child and the examiner (F of .4727 for df of 15 and 144, $p > .75$).

Essentially, then, the influence of sex, examiner and interaction variables on the weighted gain scores for Block 4 were negligible. These data were in general agreement with the results of the original study (see Table IX). The one noticeable discrepancy related to differences in gain scores by the maladjusted children under each examiner. As already indicated, examiner differences for the two groups in the present study were negligible. In contrast, the original study demonstrated significant differences between the gain scores of the maladjusted children under the two examiners since the t of 4.14 found had a probability due to chance variations alone of less than once in a thousand times (Bommarito, 1964, p. 73). Moreover, the data show that these differences were very large (see Table V). Thus, for the female-maladjusted the mean gain scores for Block 4 differed by more than 5 points (34.62 vs. 29.43); whereas the examiner differences were even greater for the male-maladjusted (39.14 vs. 31.62).

This singular finding was the only discrepancy between the two studies on sex and examiner differences where common data were compared on a quantitative basis. One statistical check was made within this study, however, which was not possible in the original investigation. In the present research, the investigator had a list of five schools which were regarded as centers of "cultural deprivation" in that they were receiving federal funds because of the large number of families in attendance with incomes of less than $3,000.00 per annum. Within these schools, the gain scores of the children under each examiner were compared with the adjustment status of each subject ignored. Thirty-four children were classified under this approach (18 for E_1, 16 for E_2). The original common mean of Block 4 for the entire sample (36.4) was then used as a critical division point to determine differences in gain scores by the children under the two examiners. If no significant differences between examiners existed, only chance variations would occur from an a priori expectation that half the scores under each examiner would exceed the common mean and half would not. The X^2 value found with correction for continuity was 5.83. With one df, such a finding could occur by chance variations less than twice in a hundred times ($p < .02$). Obviously, "culturally deprived" children performed significantly better under one examiner than under the other. For whatever value the information may have, the children under the better trained (Ph.D. vs. M.A.) examiner exhibited the lower gain scores. Significantly enough, the gain scores of maladjusted children under the more experienced examiner in the original study were also lower at a statistically significant level. But both examiners in the original study had identical training levels with Master of Arts degrees.

On the basis of the statistical analyses, it is evident that examiner differences emerged with respect to the gain scores of maladjusted children (original study) and culturally deprived students (replication). Three additional findings from both the original and the replication works complete the description of the gain scores by the Ss under each examiner. Two relate to the 1964 study and one to the present research. With respect to the former, a qualitative study of Table V indicated the following:

1. There was a steady increase in mean gain scores for all groups under both examiners over the three blocks. This evidence again indicated the significance of the amount of reinforcement.

2. Greater gain scores were made by children under examiner two in nearly all cases except for the fourth block with the female-adjusted. The advantage was very slight for the adjusted but increased somewhat for the female-adjusted. The difference in mean gain scores between the examiners seemed to be largest with the male-maladjusted, particularly in the third and fourth blocks.

With regard to the present analysis, a check was made on possible examiner differences for all four subcategories of sex-adjustment classifications (female-adjustment, female-maladjusted, male-adjusted, and male-maladjusted). As Table X indicates, none of the U values obtained reached statistically significant levels though differences in gain scores by the male-maladjusted group under each examiner had the lowest U value (p of .152). All in all, examiner differences in this study were not significant with the one exception of the cultural deprivation variable.

To review, sex and examiner differences in both studies were largely insignificant when examined from every vantage point. The two examples of examiner differences were among the maladjusted in 1964 and the culturally deprived in 1968. In the former, the differences were particularly striking, especially among the male-maladjusted pupils.

Intelligence and Gain Scores

The significance in determining the relationship between intelligence and gain scores on the learning task has three bases. First, as previously indicated, the relationship between intelligence and adjustment status may be an important one. Secondly, if the requirements of the learning task are complex enough to depend largely on problem solving behavior or high levels of verbal comprehension, the interpretation of results will be confounded. Thirdly, some recent evidence (Spence and Dunton, 1967) has shown that lower class children frequently perform more poorly than those from higher social classes in verbal conditioning tasks due to failure in comprehending requirements and not to failures in conditioning. It is for these reasons that a check on the relationship between intelligence and gain scores was deemed necessary.

With the entire sample as a basis of comparison (N of 60), a Pearson-Product Moment Correlation of .0654 indicated a negligible relationship between gain scores and I.Q. ratings on the Peabody Picture Vocabulary Scale. This coefficient was almost identical with the one found in 1964 between these two variables (see Table IX). Further, the gain scores of children with I.Q. scores of 115 or more (N of 7) were compared with those of 90 I.Q. or less (N of 6). A U test yielded a value of 13.5 with p more than .29 in a two-tailed test. Consequently, it can be concluded that no significant differences in gain scores on Block 4 occurred between groups who represented the two extremes of the sample in average I.Q. scores.

Three studies have now shown Pearson Product correlations of .07 or less on the relationship between gain scores and intellectual status (Sullivan, 1960; Bommarito, 1964, and the present study). In a word, the evidence is rather unequivocal that little or no relationship exists between I.Q. ratings and the gain scores of the learning task used in this study.

Social Class and Adjustment

In this area, there were two general operational questions investigated. First, did the children under different social classes show any significant differences in gain scores on Block 4? Secondly, did significant differences in gain scores occur among the children according to the type of position the father held? With respect to the determination of social class., the U.S. Census System (1963) as previously described was followed. Under this approach, 53 descriptions of the fathers' occupations were classifiable. In 41 cases, the father's position was matched with the same position listed in the tables and the numerical socioeconomic index associated with the job (the Prestige Score) was simply recorded. In 12 cases, this mechanical procedure was not possible and the investigator had to make 12 socioeconomic ratings on the basis of a position closely resembling the father's job.

For example, a position of salesman with industrial towels was classified as salesman--wholesale trade. In brief, 41 Prestige Scores were mechanically recorded and 12 were made on the basis of the personal judgment of the investigator. In any case, where insufficient information was provided by the teacher or the investigator experienced the slightest difficulty in classifying a position, the entry was not used. As indicated, 53 Prestige Scores were thus obtained. Since the procedure for obtaining socioeconomic indices has been explained, the data may now be discussed.

Two general findings on the relationship between socioeconomic status and gain scores on Block 4 were noted. First, a Pearson Product Correlation of .0653 was obtained between Prestige Scores and gain scores for the sample as a whole (N of 53). Secondly, the father's occupations were also classified according to the main occupational groups of the Census. In order of decreasing average Prestige Scores, these groups were (1) professional technical and kindred workers; (2) managers, officials, and proprietors, except farm;* (3) clerical, sales, and kindred workers; (4) craftsmen, foremen, and kindred workers; (5) operatives and kindred workers; (6) service workers, including household workers; and (7) laborers. U tests between the largest differences revealed no statistically significant U values for the gain scores between any two social classes (see Table X). For this reason, it may be assumed that the smaller differences would not have yielded statistically significant U values for the same gain scores.

In brief, no significant correlation was found between gain scores on Block 4 and socioeconomic indices; nor were any significant differences in gain scores obtained between comparisons of any two social classes. The final check was on the association between type of paternal occupation and the gain scores of children. As outlined previously, the finding has been that children whose fathers hold positions in the semi-skilled and service groups contribute a disproportionate number of "emotionally disturbed" children when compared to pupils with paternal occupations in the white-collar and unskilled categories. Under the present census classification of occupations, the above comparison related to differences in gain scores between occupational classes of 1, 2, 3, and 7 (N of 24) versus 5 and 6 (N of 16). A U test applied to the differences in gain scores of children with fathers in the former group versus those in the latter classification yielded a z value of .9733 (p of .1660 in a one-tail test). Accordingly, the null hypothesis of no significant differences in gain scores between the two groups on Block 4 was tenable.

---

*Not used in the present study for paired comparisons between two social classes because only one entry existed.

As a final measurement on the relationship between the adjustment classification and socioeconomic status, the Prestige Scores of children rated as maladjusted (N of 30) were compared with those of children categorized as adjusted (N of 30). A preliminary analysis on the differences in Prestige Scores revealed means of 70.68 and 60.24 for the adjusted and maladjusted groups respectively. In view of this large discrepancy in means, it is not surprising that a U test yielded a z score of 4.93 ($p < .00003$ in a one-tail test). The conclusion can be made therefore that children classified as adjusted scored significantly higher than the subjects categorized as maladjusted on Prestige Scores, the measure of socioeconomic status used in this replication study.

In sum, though no significant relationship was found between gain scores on Block 4 and socioeconomic status, a very definite and direct correspondence was obtained between SES and adjustment classifications on the teacher rating scales.

Some Clinical Findings

As the data in the original study showed, the differences on the weighted mean gain scores for Block 4, the dependent variable, showed statistically significant results in favor of the adjusted group over the maladjusted one. The results were replicated, to a partial extent at least, on the present cross-validation research. These findings are strictly quantitative. There were a number of observations which stemmed primarily from a clinical analysis of the subjects' general behavior in the testing situation and the pattern of gain scores. In a description of this analysis an expository aid may be to report the data on the original study (Bommarito, 1964) and then to describe present findings. Subsequently, an integration of data from both investigations will follow.

In 1964 study (Bommarito, 1964, pp. 82-83), the following points were particularly relevant:

1. Only two of the 60 children in the final groups failed to show gain scores between the blocks in the raw number of human preferences made. The reasons for such failure were quite different with each child. In the one case, the child had a raw number of human preferences of 25 on the first block, the maximum number possible. He simply continued to pick humans every time on all blocks. The examiner quoted him as saying: "Persons are more important than toys." He was described by the examiner as a somewhat apprehensive child who exhibited considerable compulsivity in his work habits. Finally the parents were classified as professional people whom the examiner has suspected of "pushing" the child towards achievement and the boy himself was a fairly intelligent pupil with an I.Q. of 125 on the Peabody who was classified by his teacher as adjusted. The other child was also a boy. The record for this child indicated that not a single human was chosen.

The examiner reported him to be a very frightened child who sobbed and had to have a peer with him during the testing situation. The draw a man test revealed a very primitive drawing far below expectations for the I.Q. of 95 he received on the Peabody. More significantly, the drawing exhibited a splitting of body parts, a feature, indicative of dissociation. The examiner urged the principal to refer the child to the visiting teacher immediately. Originally, the child had been classified as maladjusted.

2. Individual testing had to be discontinued with only one of the 60 children examined. In this case, the child was a girl who had successfully completed the Peabody (I.Q. 93) but who burst into "uncontrollable tears" on the fourth block of negative reinforcement. She also was referred for visiting teacher services.

3. The examiners described six children who reportedly had insight as to the purpose of the experiment. Questioning revealed that not one of them had conscious awareness or that they were not able to verbalize the "insight." Four of these children had been classified by teachers as adjusted, two as maladjusted. The I.Q. scores for all six children are in the normal range, with one score of 90 and five between 99 and 111.

4. Finally, as already suggested, the examiners referred several children for visiting teacher services on the basis of their observations in the testing situation.

In the 1968 investigation, four patterns of responses on the learning task were strikingly noteworthy.

1. Five children indicated response perseveration. In two cases the perseveration was complete in that identical responses were made on all of the 25 cards over the four blocks of reinforcement. The gain scores for these children ranged between 25-27. Four of these children had been rated as maladjusted on the teacher rating scales and one adjusted by this procedure. In the latter case, the psychologist had rated the child as maladjusted on the basis of his global clinical reaction to the child's general behavior in the testing situation.

2. Two children exhibited spatial or positional perseveration in which they tended to prefer one side of the card or the other in making human choices.[1]

---

[1] The investigator only recently became sensitized to his problem of perseveration in the responses of children to verbal conditioning (Spence and Dunton, 1967).

Both children were rated as maladjusted on the teacher rating scores and attained gain scores of 27 and 30.

3. In one case, a drastic drop in the criterion measure occurred. The case in question dropped from a gain score of 22 on Block 2 to a gain score of 7 on Block 4, with a final gain score of 26 obtained. Though his teacher judged the child as adjusted on the rating scales, she remarked that "he seeks lots of attention." Moreover, the psychologist rated him as maladjusted.

4. Three children followed a response pattern which included a spurt of increased weighted gain scores on Block 2, a drop in gain scores for Blocks 3 and 4, and initially low baseline scores. The respective gain scores for these subjects were 23, 26, and 32. The first two cases were rated as maladjusted on the basis of both teacher rating scales and clinical judgment. The last child was classified as adjusted from both procedures.

Prestige scores were available for the fathers of nine children among the group of 11 who followed one of the aforementioned patterns in gain scores. Only two of these nine children had fathers who were classifiable as white collar workers. The remaining seven fathers were in the occupational categories of skilled (N of 4), semi-skilled (N of 1), or service workers (N of 2).

To review, eleven children followed response patterns of perseveration or a drop in gain scores. Nine of these 11 children had been rated as maladjusted by both teachers and clinicians. In the remaining two children rated as adjusted, one had been judged maladjusted by the psychologist alone but the teacher herself intimated that the child was an excessive attention seeker. In effect, some basis existed for rating 10 of these 11 children as maladjusted. Significantly enough, the child rated as adjusted among these response patterns also received the highest gain score (32).

The fact that nearly all of the children following these four response patterns were rated as maladjusted and obtained very low gain scores suggested another perspective on the analysis of the data pertaining to clinical assessment. Specifically, the juxtaposition of the low gain scores and the ratings of maladjustment raised a question regarding the relationship between extremes in gain scores and adjustment status. For the purpose of this measurement, the significance levels in Fisher's exact probability statistic were used to determine whether two groups differed in the proportions with which they fell into two classifications. In the two classifications used, an attempt was made to determine what proportion of children rated as adjusted or maladjusted had gain scores one standard deviation or more above the mean (44 plus) or one standard deviation or more below (29 or less) the mean gain score on Block 4 for the entire sample. On the grounds that clinical judgment and data would enhance this analysis, children were rated as maladjusted not only on the basis of their classification on the teachers' rating scales but also by applying the following criteria:

a score of one standard deviation above or below the mean in combination with (a) a clinician's judgment of maladjustment and/or (b) behavior in the testing situation that was pathognomonic for adjustment (e.g., perseveration in response patterns, particularly 100% perseveration). The use of extreme scores on both ends of the distribution dovetails with the theoretical expectation that maladjusted children may be represented by the behavioral patterns of oversocialization or overinhibition, and undersocialization (or lack of controls and/or concern with adults). The results certainly indicated that significant differences occurred (p of .05) in the proportions with which children from the two adjustment nosologies fell within the two extreme groups (Table XI). In succinct terms, the data suggested that maladjusted children defined in the previously described manner are far more likely to show gain scores at the two extremes of the distribution than adjusted pupils.

Beyond these response patterns, the two additional findings may be briefly mentioned. First, one child evidently became satiated with the task which apparently had become an aversive stimulus to her. As she stated, "I don't want to do these anymore." This child had the worst score on the teacher scales among all the children rated with percentile ratings of P_{73}, P_{99}, and P_{99} on the Problem Checklist, the Minnesota Scale, and the Behavior Checklist respectively. Secondly, the examiners in the 1964 study listed six children who reportedly had insight regarding the reinforcement contingencies. Questioning revealed that not one of them had conscious awareness of these contingencies or that they were able to verbalize the "insight." In the present study, one child was reported to have insight regarding reinforcement contingencies.

To recapitulate, the foregoing data would suggest that the psychologist's judgment supplemented by the pattern of gain scores and clinical data yields a rich source of information for assessment purposes. In this respect, it facilitates the possible selection of overinhibited children or those lacking controls or concern about adult approval. It suggests that perseverative responses are particularly indicative of maladjustment. It implies that extremes in gain scores, especially at the lower end of the distribution, may be prognostic for maladjustive behavior. It yields a rich source of referrals to the school social worker or other school mental health personnel, sometimes to the extent of identifying children for this clinical help that teachers themselves had overlooked, even though such services were readily available. Finally, it hopefully offers a diverse body of knowledge with implications on intervention techniques and shaping procedures for improving the child's skills in behavioral coping mechanisms and general adaptation within his classroom.

VI. DISCUSSION

The most outstanding characteristic of these studies was the extreme effectiveness of mild verbal punishment in developing simple discrimination learning or concept formation. When the sample as a whole was considered, there was hardly any question that mild verbal punishment by a significant adult figure had a decided influence in shaping a child's behavior. In the original research, a definite corollary to this finding was that mild verbal punishment did discriminate between the adjusted and maladjusted, at least as determined by the definition of adjustment and the learning task criterion described in the study. In the replication, the results were somewhat more equivocal. Though the original t test indicated no statistically significant differences in gain scores, such findings did emerge with adjustments for possible procedural error in the assignment of children to the two adjustment classifications.

Several lines of evidence from both inquiries supported the powerful nature of negative reinforcement with kindergarten children: (1) the t test for human preference scores before and after conditioning: (2) an F value for treatment effects larger than that for individual differences, though the reverse is usually true; and (3) the cumulative effects over blocks. An explanation for the effectiveness of verbal negative reinforcement and the discriminatory value of the learning task criterion in differentiating the adjustment status of kindergarten children leads to the discussion of three interrelated areas: (1) the basis and evidence for the effectiveness of mild verbal punishment, (2) the implications to be made regarding its use in the socialization process, and (3) the reasons for its discriminatory function. As a final area of concern, some bases for the differences in findings among studies on verbal conditioning should be mentioned. This discussion is relevant if only because such disparity among data is common in the literature. Hence, a careful scrutiny of the reasons for these differences should be a matter of course as a prerequisite for further research in this particular area of knowledge. A return to the first topic of consideration is now in order--the basis and evidence for the effectiveness of mild verbal punishment in general.

In his review of the literature regarding the differential effectiveness not only of verbal punishment but also of the form it assumed in this study (the phrase "that's bad"), Sullivan (1960, pp. 28-31) emphasized several advantages for the technique of negative reinforcement incorporated in this research. First, the task was sufficiently simple so that mild anxiety resulting from the punishment facilitated functioning rather than impeded it. Secondly, the statement "that's bad" possibly was a stronger negative reinforcing stimulus than previously differing utterances, such as "...you are wrong." Thirdly, the punishment was performance - relevant rather than performance - irrelevant. Recent literature has provided supportive evidence not only for some of these conclusions regarding the learning task used in these studies for conditioning purposes but also for other desirable features of the reinforcement technique. The following findings are illustrative.

1. The effectiveness and superiority of negative reinforcement that is response - relevant, specific, immediate, and continuous has been well-documented (Marshall, 1965).

2. The significance of task simplicity has also been well supported (Marshall, 1965).

3. In a related matter, Spence and Dunton (1967) provided evidence for the finding that the conditioning procedure, if somewhat complex, can confound differences in performance between lower - and middle - classes, primarily because the former do not comprehend the reinforcement procedure.

In short, for various reasons, verbal conditioning tasks among young children should incorporate features of simplicity, response-relevant reinforcement, and continuous reinforcement. The verbal conditioning utilized in these studies was effective, in part, because it incorporated these very characteristics.

As previously indicated, the age of the child may have a strong bearing on the effectiveness of mild verbal punishment for at least two reasons: (1) the avoidance of anxiety may have served as a strong negative reinforcer in the light of reinforcement history; (2) as the study by Sears, Maccoby, and Levin (1957) suggested, in the early years of socialization the ratio of punishment to rewards may be in favor of the former. In any event, this investigation clearly indicated that mild verbal punishment is very likely to be used prevalently by parents in the socialization of the child. Thus, by a process of stimulus generalization the child who has had a supportive relationship with the parents will want and/or expect the same relationship with the significant school adult. Accordingly, just as mild verbal punishment by the parent will arouse anxiety in the child so will the same behavior by the teacher or a teacher-figure elicit a similar response in him. In either case, the result is the same; the child becomes susceptible to adult influence because he wants to please adults and maintain a pleasurable supportive relationship. He thus wants to avoid the anxiety that results from a disturbance of this relationship. It is no wonder, then, that research on discrimination learning among kindergarten children has found mild verbal punishment to be an extremely effective reinforcer in its own right, generally superior to positive reinforcement alone, and almost equal in effectiveness to the combination of reward-punishment (Kelly, 1966; Marshall, 1965; Spence and Dunton, 1967).

It is true, no doubt, that punishment could have serious drawbacks as a means of developing socialization. First, it could develop a generalized avoidance reaction to the situation from which it arose. Accordingly, the child may learn to avoid significant adult figures, an undesirable consequence. Secondly, it could be too effective in that rigidity of behavior could occur (Sullivan, 1960, p. 36).

However, these undesirable consequences may not occur wheremild verbal punishment is applied in a situation where the child's alternatives to action or choice are clear and specific, as illustrated by the learning task criterion in this study. The implication is, however, that they may occur where the expectations for behavior by the adult are ambiguous for the child. It follows, then, that under the proper circumstances, mild verbal punishment as a means of socialization may have its place. Indeed, both Sears and his colleagues emphasized "...for maximum effectiveness the actions that the mother rewards or punishes should be distinct and specific," (1957, p. 360). In the writer's opinion, the words "significant adult figure" could have just as easily been substituted for mother in the above sentence with little or no loss in meaning. At any rate, recent research has provided considerable suggestive evidence in support of the foregoing conclusions by Sears and his cohorts (Marshall, 1965).

At this point, the investigator should like to emphasize that he does not want to become involved in the controversy regarding the relative merits of reward versus punishment in the socialization process. He would argue that both may be effective under the proper conditions. For scholarly discussions on the merits stemming from the judicious use of punishment, the interested reader should consult the literature (e.g., Bandura, 1962, Michael and Myerson, 1965).

The crucial conclusion of this paper, however, does not rest with the relative merits of reward or punishment but rather with the discriminatory function of negative reinforcement to determine adjustment basis through concurrent validation.[1] To review, the assumption is that it is a function of the reinforcement history of the child. This history, in turn, conceivably has a number of interrelated major characteristics which lend support to this assumption. First, negative reinforcement is prevalently used in our culture, a point the study by Sears emphasized. Secondly, one reason it is so prevalent is because it is almost unavoidable. Sear's study also supported this conclusion. Another example of the almost inevitable occurence of negative reinforcement for the child is the kindergarten room, or any school room for that matter. In this situation, a child has to learn to comply to certain rules and regulations. Because of the need for orderly proceedings of some sort even in the most loosely structured school activity, the teacher does not hesitate to apply performance relevant punishment or reward. But these two factors are not sufficient reasons to explain the effectiveness or significance of mild verbal punishment in predicting adjustment status. A very plausible third factor is that both rewarding and punishing behavior by the parents have become generalized reinforcers for the child. By a process of stimulus generalization, reward and punishment by a significant adult outside the home have become generalized reinforcers for a child's behavior. By way of explanation, the parents with fairly adequate relationships with the child frequently provide loving care and satisfaction for needs of many kinds. The stimuli arising from parental attention become learned or conditioned generalized reinforcers for the child independent of his condition of deprivation at any given moment (Holland and Skinner, 1961, p. 71). In like manner, words like "bad" and "wrong" are frequently heard just before punishment. These words come to elicit responses characteristic of anxiety. They also become learned or conditioned stimuli (Holland and Skinner, 1961, p. 33). Similarly, as already indicated, these words are most likely to be frequently employed in the socialization process with young children by the parents of our culture.

---

[1] Some social workers and educators have criticized the present research because it engenders anxiety in the young child, a matter which they regard as an "ethical" consideration. By this comment, they display a gross misunderstanding regarding the purpose of the research. Specifically, if a child "shatters" in the face of the mild punishment or does not respond to it, his behavior presages poorly for his adjustment in school since it indicates that the principal behavioral control mechanism of the teacher--verbal reinforcement--will not be reinforcing to the child or that it will be overwhelming to him. In either case, he's going to need considerable help for his social adaptation. The argument by these critics resembles the old "strawman" on the "evils" of behavioral control. Needless to say, the investigaor has yet to meet a reputable learning theorist who has offered such a criticism.

To review, by a conditioning or learning process, reward and punishment have become reinforcers for the child, independent of his deprivation at any given moment; that is, they have become generalized reinforcers. As explained before, mild verbal punishment as a generalized reinforcer is likely to be much more effective with children who have had strong positive reinforcement histories with their parents; such children have "something to lose" if their parents withdraw support. Under these circumstances, it is the desire of the child to avoid the anxiety caused by the possible loss of this satisfactory relationship to a significant adult figure that renders his behavior susceptible to change by the latter. In the absence of this susceptibility, poor reinforcement histories may be assumed and maladaptive behavior inferred. Moreover, it is highly likely that in a classroom environment where the situations make some order mandatory, the child whose behavior is not susceptible to influence by a teacher will come to be in conflict with her. He will, in short, exhibit maladaptive behavior, and for all practical purposes, may be considered maladjusted. In a word, this inquiry has suggested that mild verbal punishment as a means of eliciting changes in behavior should be much more effective with relatively adjusted children than maladjusted ones.

Thusfar, the discussion has centered on the basis and evidence for the effectiveness of mild verbal punishment, the implications to be made regarding its use in the socialization process, and the reasons for its discriminating function. Despite this impressive evidence regarding the effectiveness of mild verbal punishment in shaping children's behavior, the research data indicate considerable confusion and conflict among the findings (Kennedy and Willcutt, 1964; Marshall, 1965; and Spence and Dunton, 1967). The data also show contradictions in findings with replications by the same authors (e.g., Spence and Dunton, 1967). It is quite appropriate to examine the bases for these inconsistencies if only to provide clarification for future directions in research. What follows are illustrative but by no means exhaustive indications of contradictory data.

In their review of the literature relating verbal praise and blame to performance, Kennedy and Willcutt (1964) found that stress (engendered by adult disapproval) facilitated the performance of low-scholastic aptitude subjects; yet they also cited studies indicating that verbal praise, rather than stress, had a facilitating effect on the performance of mentally defective children who are notorious for their low scholastic aptitude. Similarly, the major finding by Spence and Dunton (1967) that the combination of verbal reward - punishment was a more effective reinforcer than a material reward (candy) contradicted findings by others showing that material rewards were superior for lower-class children (Terrell, Durkin, and Wiesley, 1959; Zigler and DeLabry, 1962). As a final example, the principal conclusion in the research by Zigler and Kanzer (1962) was that verbal praise ("good, fine") was a more effective reinforcer with lower-class than with middle-class children, while the effectiveness of "correct" ("right, correct") reinforcers operated in reverse fashion. But Spence and Dunton have attributed such differences not to the comparative effectiveness of reinforcers but to a lack of verbal comprehension on the part of lower-class children concerning the reinforcement procedure under "right-blank." Obviously, confusion is commonly found among the research data on the use of verbal reinforcement with children.

One obvious basis for this confusion is the sheer number of studies now available in the research literature on verbal conditioning. As Krasner (1965, p. 214) has emphasized, the list has grown from 35 studies 10 years ago to several hundred for the present time. In his opinion, the literature is so vast that it is even impossible to tabulate proportions of positive or negative results. Moreover, as he has cautioned, such a comparison would be meaningless. "A 'negative' study, at this point, usually means that the investigator's hypotheses were not confirmed. Conditioning may have occurred, but the special relationships hypothesized were not." In effect, then, Krasner is saying that the evidence supporting the occurrence of conditioning is quite decisive.

Beyond the factor concerning the unmanageability of massive data, Marshall (1965, pp. 27-32) cited and elaborated considerable evidence indicating that the effectiveness of mild verbal punishment as a reinforcer for learning relates to task complexity, performance-relevant reinforcement, and the experimenter. He especially stressed the meager or complete absence of control over the variables of achievement orientation, personality traits, ego involvement, and level of aspiration among subjects. To this list, the crucial variable of reinforcement history should be added. Kennedy and Willcutt (1964), for example, reviewed 33 articles spanning 50 years of research relating to the effects of praise and punishment on the performance of school children. On the basis of this painstaking inquiry, they concluded that the problem of subject reinforcement history was the major contributor to the confusion regarding the effects of reinforcement on the performance of children.

Doubtless, many vital and complex variables affect the performance of children under conditions of verbal reinforcement. Manifestly, these variables must relate to the gain scores of the children in the present studies and to the general value of using conditionality as a screening instrument for emotional disturbance. By and large, the majority of these variables will be incorporated in the remaining discussion under the categories of (1) examiner differences, (2) mental age, (3) social class status, (4) the present usefulness of the screening instrument, and (5) suggestions for further research.

Examiner Differences

In the 1964 study by the investigator, maladjusted children under one examiner obtained far greater criterion scores than subjects under the other psychologist. These differences were very large and particularly applied to the gain scores of maladjusted boys. In this research, the psychologists were females widely disparate in chronological age and professional experience. In the present research, examiner differences were noted only for "culturally deprived" children. The examiners, in this case, were males with sharp differences in professional training, moderate dissimilarities in chronological age, and approximately equal in clinical experience. In the original research (Sullivan, 1960) which in essence, served as a pilot study for the present investigations, no examiner differences whatsoever emerged. The two male examiners were equivalent, moreover, in almost all essential characteristics--age, level of training and experience, familiarity with the task, research knowledge, and even sophistication with learning theory. In any event, it is quite apparent that the results relating to the examiner variable were hardly consistent. How can these inconsistencies be explained?

An immediate answer to this question should be readily apparent. These differences are to be expected when one considers the significance and complexity of the examiner variable in psychological control, as well as the difficulty in obtaining experimental control over the examiners' behavior. Sarason (1965), who has devoted considerable research to this topic, has summarized most of its essentials clearly. It is his belief that the experimenter variable is probably the most neglected element today in psychological research. Accordingly, he questions a large percentage of the research with human subjects because of unmeasured examiner effects. Among the many critical ingredients of the examiner, he describes the personality correlates of the experimenter and the possibility of changed behavior by him in the course of an experiment as two very vital elements requiring control. He describes research which shows, for example, that responses by Ss in a conditioning task have been related to the degree of dominance held by Es as inferred from personality scales. In closely related areas, it has been found that the attitudes of Es towards Ss and the expectancy sets of the former for performance influences the results. Again, it has been found that the behavior of the experimenter may undergo change with added trials or practice effects. As Sarason stresses, the possibility of significant behavioral change by the E (e.g., reduced anxiety) with increased number of trials may be one major source for the unreliability of many types of psychological research.

To sum up, Sarason has recommended controls for subjects' and experimenters' personalities, sex differences, and other individual differences. Additionally, he has suggested controls for the following variables, whether dependent or independent, in psychological research.

1. The degree of psychological sophistication of the experimenter and his level of training in research.

2. The skill in adaptation by both E and S to the social situation in which they are functioning. This adaptation, in turn, relates to the length of the experimental session, the intervals between sessions, and even the time of day in which the sessions are held.

3. The existence of prior contacts between S and E, as well as the number of contacts.

4. The friendliness of the E and S to each other in the testing situation.

5. Finally, Sarason has posed the following relevant question: "To what extent, do subject, experimenter, and situational variables have comparable effects over tasks. It seems possible that some experimenters [or teachers] may be better at eliciting certain response classes from certain types of subjects than from other subjects."

One final comment regarding the significance of the examiner influence on results deserves careful scrutiny. Reference here is to the prestige of a specific examiner with a given child. Comprehensive reviews of research concerning the experimental manipulation of verbal behavior have shown that the effects of similar verbal reinforcement varies directly with the prestige of the reinforcer. The prestige itself, in turn, depends upon the S's reinforcement history with a given examiner (Salzinger, 1959). As McCandless (1961, pp. 130-132) has explained, moreover, through the process of learned or secondary stimulus generalization from reinforcement histories different adults could conceivably have different effects on children. Recent research evidence has also supported this conclusion (Patterson, 1965). It is clear, then, that variations in the prestige of the examiner may occur as a result of reinforcement histories of the child in general and with a given examiner in particular.

The applications and implications of examiner prestige were evident in the research by Sullivan (1960), Bommarito (1964), as well as the present inquiry. The latter, for example, was the only instance in which both examiners had no prior contact with the children tested. Regardless of this difference among the studies, the implication for further research on conditioning among children is distinct; namely, the necessity of control, whenever possible, for the prestige of the examiner.

On the basis of the evidence, there is no question that control and measurement of the examiner variable is quite essential and important. Yet, such control is extremely difficult to obtain, a point which Bijou and Baer (1963, p. 216) have aptly clarified: "But a person is a difficult dispenser to bring under tight experimental control; he is himself sensitive to the behavior of the child with whom he interacts, and may respond to various acts of the child in ways which provide (or remove) extraneous or incorrect social reinforcers. A nod, a smile (even a smile suppressed at the expense of quivering lips), lifted eyebrows, sudden glances--any or all of these (and other) reactions of the experimenter may serve to reinforce the child, and ruin a critical experimental contingency." In sum, the experimenter may have unawareness for the existence, intensity, nature, and frequency of the reinforcers he dispenses in his interaction with children during an experimental session.

Mental Age

As previously documented, three studies over an eight year period have now indicated that performance on the learning task utilized in the present inquiry is negligibly related to intellectual level. Nevertheless, this variable needs careful consideration in future research on verbal conditioning for several reasons. First, some comprehensive longitudinal research (Anderson, 1959) has found the prediction of social adjustment inextricably interrelated with intelligence. It follows that, if verbal conditioning is to predict adjustment status, the intelligence of the child cannot be completely ignored. Secondly, failure to control this variable may lead to confounding interpretation of results. For illustration, Spence and Dunton (1967) concluded that children in lower socio-economic classes (mostly Negroes) performed more poorly than subjects from middle classes under the reinforcement contingency of "right-blank" because of failure to comprehend the reinforcement procedure. The question immediately arises as to whether these differences were attributable to class memberships or intellectual differences. In other words, would these class differences have held with intellectual levels constant? Since the authors provided no data whatsoever for this crucial variable, the question must remain unanswered. Finally, the lack of a relationship between intelligence and verbal conditioning holds for the simple discrimination task utilized for the three studies under consideration. If the complexity of the learning task increases, this finding may no longer apply. As one example, an experimental session involving free operant conditioning would present a much more complex task to a child.

Under these conditions, the influence of intelligence on gain scores remains a vital matter for investigation.

Evidently, then, the continued concern over the interrelationship between intelligence and gain scores rests on at least three premises: (1) intelligence has been associated with the prediction of adjustment over time; (2) the variable may provide an alternative explanation to differences in the effectiveness of divergent reinforcement procedures; and (3) the factor of intellectual level would assume even more significance with an increase in task complexity.

Social Class Status

The results on the relationship between gain scores and social class status have hardly been consistent and unequivocal. On the basis of the combined data from the original study and its replication, the singular finding emerging was that a relationship existed between the criterion score and the type of paternal occupation of the father (Bommarito, 1964). The present investigation did not cross-validate this datum. It did, however, show a statistically significant difference in levels of socioeconomic status (SES) between adjustment groups in favor of adjusted children (as classified by the teacher rating scales). The divergency in data may stem from many sources: (1) the inadequacy of the indices for measuring SES, (2) the change in the SES indices used between the 1964 study and the present one; (3) inadequacies in sampling procedures; (4) the insensitivity of conditionality for the prediction of personality differences; and (5) the questionable value of the conclusion that one should expect to find a higher incidence of maladaptive behavior among children from lower-social classes than from middle-classes.

In the investigator's estimation, the sampling procedure is the factor most vulnerable to scholarly criticism. It would be no difficult matter, however, to find supporters of the other objections. In rebuttal, two counterarguments need to be offered immediately. First, there is no question that the five problems mentioned above deserve careful consideration in any further research on verbal conditioning. Each situation is so complex, in fact, that it deserves separate treatment. For these reasons, these matters will not be discussed for the present. Secondly, the investigator is unable to entertain the notion that no differences in adaptation problems exist between children from different social classes. The sum total of the evidence previously described would negate the tenability of this notion. Nevertheless, he would question the value of the SES variable per se as a predictor of a behavior. If only because sociologists place so much emphasis on SES as a causative agent of behavior, a challenge to this cherished belief requires a detailed explanation.

SES is a global and non-specific variable. As such, it does not contribute much to a functional analysis of behavior. A more pragmatic approach might be to consider specific components which are detrimental to learning or to the formation of desirable social behavior. To the extent that certain specific but determinable factors are influencing a person's skill development, impairment will follow. Though the chances are that children from lower SES groups will have a greater number of these growth retarding elements in their environment, this frame of reference may account for within group differences also. It is this approach which Whiteman and Deutsch (1968) have adopted and which has led to the formulation of a social deprivation index. As they have stated, "Social deprivation implies that the association between social grouping and specific environmental factors is not directly causal but is mediated by more basic societal conditions as unemployment, poverty, and inequality of opportunity in various areas. With the removal of such conditions, the association between social grouping and disadvantaging factors may vanish."

These authors have described several functions that this objective index of social deprivation may serve, including the role of specifying detrimental environmental conditions, the role of mediating the relationship between SES and test scores, the role of an independent contributor to behavior beyond the independent variables of SES and race, and the role of interactive variable. These investigators have presented specific data to support the validity of the instrument for these various roles. They have also described the precise nature of the index in full.

In the investigator's estimation, the use of this index is preferable to the application of SES per se because of its many potential applications for a functional analysis of behavior. In terms of learning theory, for example, the utilization of the index dovetails closely with the three broad questions professional workers interested in behavior modification are likely to ask: "(a) what behavior is maladaptive, that is, what subject behaviors should be increased or decreased; (b) what environmental contingencies currently support the subject's behavior (either to maintain his undesirable behavior or to reduce the likelihood of his performing a more adaptive response; and (c) what environmental changes, usually reinforcing stimuli, may be manipulated to alter the subject's behavior" (Ullmann and Krasner, 1965, pp. 1-2). In other words, the data procured from the utilization of the deprivation index provides many possibilities for responses to these crucial questions in behavior modification.

The investigator's disenchantment with the inclusion of the SES variable in conditionality studies does not rest solely with its failure in contributing meaningfully to a functional analysis of behavior. Another basis is the matter of preferential priority. Conceivably, other social-psychological correlates of mental health may be of more significance to scrutinize than SES per se. As illustrations of these frequently neglected correlates, one could include urban-rural differences, intracity differences as between dwelling in inner city tracts versus peripheral ones, racial and ethnic differences, role incompatability between behavior learned in one's primary culture and contrasting expectations from an immediate but long-standing and continuous social situation (e.g., the school), and the sudden loss of affectively charged relations (Scott, 1958a).

With respect to these social psychological correlates, it is also quite conceivable that the crucial variables on the relationship between social class status and adjustment as defined in these studies rest on entirely different grounds than social membership alone. Perhaps the most pertinent variable is suggested from the very thorough study by Bettleheim and Janowitz (1950). In this work, they convincingly demonstrated the devastating adverse effects of downward social mobility on the personality of adults. It follows, then, that the downward social mobility of their parents may affect the adjustment of children. In a related area, under certain conditions, the relationship between an upwardly mobile family and gain scores would be worthy of consideration. Swanson (1961, p. 21) has presented this point very forcefully: "social scientists tend to think that the upwardly mobile who advance through their own efforts [the entreprenurial middle class families] are disproportionately likely to leave no effort untried in expunging behaviors and desires which might impede their ascent...." Similarly, "occupations in which competition is exceptionally severe and judgment of attainment highly impersonal (e.g., bureaucracy) will be marked by a high incidence of mental disorder...." (Scott, 1958, p. 831). The implication from these researchers is that social-psychological variables other than social class membership conceivably are more important for a child's adjustment status. The variable of downward mobility by Betteheim and Janowitz might be particularly promising with respect to children's adjustment in view of the lack of data here. To this worker's knowledge, no research at all has been reported on this relationship.

To summarize, the data on the relationship between SES membership and gain scores have not been consistent. Though statistical methodology, instrumentology, and certain challengable premises may all be contributing to this inconsistency, an alternative two-fold approach was described which may yield more productive data than the SES of the subjects. On the one hand, a recently developed deprivational index was recommended as dovetailing closely with a functional analysis of behavior as promulgated by behavioristically oriented professional workers. On the other hand, the study of other social psychological correlates of mental health possibly of more significance for a child's adjustment status than SES per se was advocated. These were described in full.

The Present Status of the Screening Instrument

"When one's goal is purely theoretical ... broad group differences constitute sufficient evidence for the verification of assumptions" (Kinget, 1952, p. 5). This statement aptly summarizes the present value of the learning task criterion as an instrument for screening maladjusted children. The evidence appears to have substantiated, to some degree, the assumption that adjustment can be predicted by susceptibility to negative reinforcement. The additional assumption that variations in this susceptibility are a function of differences in the reinforcement histories of individuals has yet to be verified. If the screening instrument is to be a useful one to the schools, however, several minimal improvements must occur as suggested below.

1. The replication of results needs to be on a firmer foundation of data than now exists principally because the cross-validating research provided evidence on an ad hoc basis. The prediction of outcomes would provide the needed statistical support. For this purpose, helpful procedures would include the elimination of procedural errors, alternative choices of external validating instruments, a larger sample, and the utilization of a different basis for the selection of the original sample other than a simple random design.

2. Reliability of the instrument needs to be established in some manner. One possible procedure would be to use equivalent forms with equated groups. Such a method would have the added advantage of verifying the premise that it is the reinforcement procedure which influences behavior, not the nature of the stimulus cards.

3. Predictive validity would undoubtedly enhance the value of the instrument. To this end, gain scores young children receive prior to any schooling would be compared with teacher's ratings and also be used to predict these evaluations of behavior. Changes in teacher's ratings over time could also be compared with changes in gain scores over time.

4. The relationship of gain scores to the type of adjustment by the S (e.g., whether overinhibited or undersocialized) needs further exploration. The present study has verified this relationship to some degree but the issue needs more explanation in depth because of its implications for intervention techniques in the shaping of children's behavior. This may require a search for more appropriate scales than the present scales which appear to be rather unsatisfactory for this intention.

Additional data on cross-validation and more information on concurrent validity, predictive validity, and reliability--these are the minimal requirements needed to establish the practical value of the learning task as a screening process for maladaptive behavior among children.

Directions for Further Research

In studies on verbal conditioning, Sarason (1965) has emphasized the need for controls over three sets of variables: (1) individual differences among subjects, for example, personality variables and sex; (2) individual differences among experimenters; and (3) social, interpersonal, and situational variables, for example, the way in which the subject's task is structured to him. Since these sets of variables encompass the essentials of research on verbal conditioning, future directions for the present research have been subsumed under each area for both immediate applications and long-range extensions.

In terms of the more immediate applications, a careful study needs to be made with respect to (1) the adequacy of the present conditioning task, (2) the conditionality of personality types, (3) task structure, (4) examiner control and (5) methodological considerations.

Adequacy of learning task. The adequacy of the learning task essentially refers to two evaluative dimensions--its appropriateness as a screening instrument and the desirability of negative reinforcement as a measure of conditionality. The first dimension has already been discussed under the section entitled the present status of the screening instrument. The second aspect of the task requires slightly more attention. Three valid generalizations apply to this consideration. First, the task meets desirable specifications for conditioning among young children in that it includes the characteristics of simplicity, performance-relevant reinforcement, and continuous reinforcement. Secondly, as already shown, negative reinforcement has been proven to be very effective with young children, being superior to positive reinforcement alone and almost equally powerful as the reinforcement combination of positive and negative reinforcement. Finally, (Marshall, 1965, pp. 31-32) provides evidence indicating that much of the confusion among the data from reinforcement studies disappears if performance relevant and nonspecific performance reinforcement are separated. It should be clear that the present learning task is an adequate instrument for the measurement of conditioning.

Conditionality of personality types. The essence of the present learning task, of course, is that it can predict adjustment status by differences in responsiveness to negative reinforcement. The value of the learning task as a screening instrument will depend, in part, on the efficiency with which it can make such predictions. The theoretical basis for this expectation has already been treated comprehensively. Generally, the major assumption was that adjusted children are more responsive to social influences as embodied in the learning task because of reinforcement histories in which they have learned that adults are dependable and a source of satisfaction. Basically, they have learned to please adults because it is pleasant to do so and unpleasant (anxiety-arousing) not to do so. A subsidiary and related hypothesis is that personality types or certain behavioral tendencies may also be related

2

to conditionality. Within this context, the following relationships will need to be carefully investigated in the interest of improving the screening instrument.

From a review of the research by Bandua and Walters (1963, pp. 201-203), it is possible to conclude that responsiveness to conditioning is heightened among neurotic types and reduced among anti-social types. A closely correlated finding is that social reinforcers are very likely to be effective with children who have learned strong dependency needs (Bandua and Walters, 1963, p. 10). Factor analysis has also found the two behavioral dimensions of psychopathy, or lack of controls, and neuroticism, or overinhibition (e.g., see Appendix A). Perhaps Patterson's conclusion (1965, p. 177) stemming from his own work and a review of the research literature, gives additional depth to this matter of conditionality. In his estimation maladjusted children are likely to be characterized by a lack of responsiveness to social approval or an excess of responsiveness to social disapproval.

In review, psychopathy may be associated with reduced responsiveness to negative and/or positive reinforcement, neuroticism by overresponsiveness to these reinforcement contingencies. To some extent, the present research verified the relationship between extremes in gain scores and maladjustment, principally by combining results on the learning task with test behavior (notably perseveration) and clinical judgment. However, the relationship between conditionality and personality type needs a more systematic inquiry than was possible in the present research and possibly improved external validating criteria for the evaluation of behavior.

Task structure. The nature of the learning task in its present form already has much that is commendable. Nevertheless, it may be highly informative to compare the effectiveness of the learning task as it stands now with three variations; namely, a free operant learning task, careful instructions to the S on reinforcement procedures, and extinction. The advantages of each variation need brief elucidation.

Krasner (1965) has suggested that free operant learning tasks replace those requiring limited response repertoires in verbal conditioning studies because of the lack of response generalization in these latter situations. For example, singularly, or in conjunction with his colleagues, Krasner presents projective type cards to which subjects are requested to respond in the form of a story. Reinforcement then follows all emotional words or some variation of this procedure (e.g., reinforcement of only pleasant emotional words). He has cautioned, however, that the reinforcement procedure should also be explained explicitly to the examiner(s) prior to any experiment in terms of reinforcement theory, operational definitions, and precautions on behavioral cues by the E providing the S reinforcement unwittingly. The classes of behavior to be reinforced should be clearly defined to the E but the specific manner of reinforcement should be left to the E to do what is natural for him within the limitations prescribed by the research. Beyond these research precautions, Krasner has also stressed that free operant task should also include such characteristics as a disguised form, face validity, and the appearance of a plausible research undertaking.

The advantages of this procedure are reportedly numerous. First, it eliminates efforts on the part of the S to determine examiner expectancies for task performance. Secondly, it reduces the chances of awareness by the S regarding reinforcement contingencies because of the complex relationship between variables. Thirdly, it resembles "real" life and other behavioral influences, such as psychotherapy, more closely. Lastly, it is probably more resistant to extinction than one-trial learning in which reinforcement contingencies are made explicit to the S prior to any task presentation. Manifestly, the procedure presents a rich potential source of data for conditioning studies.

In spite of Krasner's opposition to explicit explanations regarding reinforcement contingencies, the procedure presents an additional variation of the task structure possibly worthy of trial. At least two basic advantages accrue to this variation in the reinforcement procedure. First, it possibly eliminates or reduces misinterpretation and misunderstanding among children regarding the reinforcement procedure (Spence and Dunton, 1967). Secondly, it serves as a test of Rotter's social learning theory (1954) which affirms that the probability of occurrence of a given behavior in a particular situation depends on two determinants--the subjective expectancy by the S that the behavior in question will be reinforced and the value of the reinforcer to the S. Certainly, explicit directions on the reinforcement procedure prior to the experiment should strengthen the subjective expectancy of reinforcement. Furthermore, if adjusted children have a greater desire to please adults than maladjusted ones do, this procedure enhances the opportunities for expressing this desire since expectancies for behavior from the child by the adult will be clarified. For these reasons, the procedure deserves recognition as a possible improvement in the learning task.

The final variation suggested for the task structure is the inclusion of extinction. In this connection, Salzinger and Portnoy (1964), using a verbal conditioning task in a free operant situation with acute schizophrenic adult patients, found that the prediction for release from the hospital 180 days later was far superior for patients who conditioned and extinguished than it was for those who did not condition or who conditioned but did not extinguish.

The addition of these three variations to the task--free operant conditioning, careful instructions on the reinforcement procedure, and extinction--offer a large number of variations in the reinforcement procedure for exploration purposes. An illustrative list of these possible variations follows: (1) the present task, (2) the present task with extinction, (3) the present task with prior knowledge of reinforcement procedures made explicit to the child, (4) free operant conditioning, and (5) free operant conditioning with extinction. Furthermore, with the use of positive and negative verbal reinforcement individually or in combination, this list of variations easily doubles.

Examiner differences. As already elaborated, the examiner variable in psychological research, though crucial, not only is frequently overlooked but also is extremely difficult to control. Nevertheless, because it is so significant, it requires such control. The solution to this problem is to train the experimenter to respond to the child only according to experimental specifications or to use some type of apparatus as standardized stimuli for behavior (Bijou and Baer, 1963, p. 216). Sarason (1965, p. 237) presents some trenchant reasons for not favoring this latter approach, one of which is that it definitely does not eliminate the experimenter variable. In any event, a field-experimental study at the present would find it neither feasible nor desirable to eliminate this variable. Without necessarily covering all or even a majority of the dimensions in the experimenter variable, research on verbal conditioning in the near future could assume several forms. First, the examiner(s) could be instructed and trained to respond to the child only according to experimental specifications. Direct observations of "trial runs" may add additional assurance in adherence to these specifications. Secondly, insofar as possible, the examiners should be matched on objective variables such as age, sophistication with learning theory and research, prior contacts with the Ss, etc. Finally, under certain conditions, it may be desirable to measure differences between examiners rather than eliminating them. Some evidence exists, for example, that the performance of Negro children is superior with Negro examiners than with white adults. This superior performance has been demonstrated in discrimination learning under blame (Kennedy and Willcutt, 1964) and in general test results (Baugham and Dahlstrom, 1968). Race of the child as it interacts with race of the examiner, then, comprises a fruitful area of investigation.[1]

Methodological considerations. As mentioned and described before in complete detail, the present study very possibly was characterized by a procedural error. The verification of the major hypothesis after statistical adjustments for this possible error would lend substance to this assertion. This phenomenon would suggest that methodological considerations in future studies require careful thought. Among the many possibilities here, three crucial issues are the assignment of the children to adjustment groups, the sampling procedure, and the sampling size. The recommendations for these matters by the investigator would be as follows. First, whether the children are classified within an extreme adjustment group on the basis of raw scores or normalized standard scores is not too crucial. More significant perhaps is that the classification occur on the basis of extreme percentile rankings on all scales. Secondly, the method of random sampling has a potentially devastating drawback in that it may select the most disturbed children within a class on a chance basis. Several teachers commented on the fact that the investigator had failed to select their most poorly adjusted children.

---

[1] Since the present study included only three Negro children, the measurement of this interaction was not feasible.

One method that would circumvent this disadvantage would be to require teachers to select the most poorly adjusted children within her class. The screening instrument by Bower and Lambert (1961, 1961a) lends itself to this purpose admirably. Finally, there is no question that a significant increase in the sample size for the original pool of children to be rated by teachers would increase the reliability of the rating scales and that very large samples are needed for appropriate reliability (Arkin and Colton, 1963). The same considerations apply, of course, with regard to the need for a larger sample in the conditioning task. The classification procedure, the sampling method, and the sample size--these are minimal methodological considerations needing attention.

Now that an inventory of more immediate applications has been completed it is appropriate to establish a preferential priority list of desirable sequences. In this capacity, the sequential preferences of the present investigator would be (1) a repetition of the study with due attention to the methodological issues just outlined, even if it were necessary to use the same teacher scales; (2) a search for more adequate external validating criterion; particularly for instrumentology providing valid and reliable discrimination of personality types; and (3) the simultaneous use of several variations in the task structure. Inherent in all of these sequences would be careful control over the examiner variable. Subsequenly, further extensions may be attempted. It is to these possible long-range extensions in the use of the learning task that will now be discussed.

The long-range extensions may be placed on a time continuum of need. Eventually, the relationship between responsiveness to conditioning on the one hand and deprivational indices and other social psyhological on the other will require checking. On an intermediate level of urgency the nature of the association between subject and examiner attributes and conditionality merits appraisal. On the subject side, the S's level of aspiration, achievement orientation, and ego involvement in the task requires some research effort, particularly as these matters relate to the prediction of adjustment skills. On the examiner variable, the influence of his personality, as it is inferred from personality scales, on children's conditioning provides a worthy avenue of inquiry. In the investigator's judgment, however, the most pressing needs for the extension of the screening instrument pertain to (1) school and classroom aids; (2) reinforcement histories; and (3) schedules of reinforcement.

School and classroom aids. Extensions in the application of the learning task might aid teachers and school personnel in the following ways.

1. The instrument might be used to measure the effectiveness of a therapeutic environment (for example, special class for the emotionally

disturbed). Thus, if a child is unable to be conditioned to the learning, whatever the reason may be, it is logical to conclude that if the therapeutic environment were effective one of the results should be that the child is now able to be conditioned for such learning. If the changes in conditionality occurred for both the teacher and some outside ancillary personnel, such as a school psychologist, generalization of improved behavior would be a safe assumption to make. With such generalization evident, release from the special class would rest on firmer footing than would have been otherwise possible.

2. Krasner (1965, pp. 223-226) has presented evidence from four separate but related comprehensive studies which buttress his conclusions that "verbal conditioning is a major aspect of traditional psychotherapy; verbal conditioning is prototypical of the social scene; verbal conditioning is a treatment procedure in its own right." In like manner, Salzinger and Poutroy (1964, p. 9) have asserted that "reinforcement theory in conjunction with a controlled interview, i.e., one in which reinforcement contingencies are predictable and precise, serves as a useful technique for diagnostic and therapeutic purposes." Ancillary personnel in the schools, such as school psychologists, and classroom teachers certainly should be able to derive many benefits from this technique. For example, it should be apparent that responsiveness to verbal conditioning might serve as a useful prognosticator for movement in therapy. Again, it is a well-established principle that very few social reinforcers are effective with psychotics (Krasner and Ullmann, 1965; Ullmann and Krasner, 1965). The degree, intensity, or absence of responsiveness from children during conditioning tasks would not only provide clues on the severity of a child's behavioral pathology but also data regarding the nature of his difficulty and the reinforcement or shaping procedures which may be most helpful. Such an approach would even have implications for the type of setting in which the treatment should take place along the possible bi-polar dimensions of special-class vs. regular class, tutorial arrangement vs. small group, institutional setting vs. public schools, and parental treatment vs. professional treatment. With respect to this last bi-polar dimension, Patterson (1965), for one, has stressed that under simple paradigms of reinforcement contingencies parents can be taught to shape the child's behavior, if the deviancy is not too flagrant, particularly among young children. His own studies have demonstrated this principle admirably.

3. The significance of modeling in shaping behavior, particularly when combined with reinforcement, has been well established (Bandura and Walters, 1963). But the effectiveness of modeling can also be demonstrated to teachers as part of a simple conditioning task (Kelly, 1966). The most desirable end, of course, would be for the teacher to incorporate implications of modeling in her management of individual or group behavior.

4. Patterson (1965, p. 176) perceives the responsiveness to social stimuli (RSS) variable as composed of three distinct but overlapping components: responsiveness to social stimuli as approval and disapproval, imitation, and ability to discriminate among social cues. Ideally, the scope of the learning task's complexity should be broadened to incorporate these components. With this added versatility and breadth for sampling the adaptive skills of a child, the conditioning task would become more efficient in selecting vulnerable children who need help and in suggesting the kind of supportive aid from which they could most profit. Even in its present simplistic form, the learning task could be employed by kindergarten teachers as a clinical tool rich in potentiality for data. As one example, the notation of highly perseverative behavior needs clinical study. In the present research, all children with such behavioral tendencies were rated as maladjusted. Additional evidence on the significance of this behavior is provided by Spence and Dunton (1967) who found that children who perseverated continued to do so even after specific instructions designed to eliminate it. The accumulation of evidence from these two studies would suggest that perseverative tendencies are associated with lowered responsiveness to social reinforcement. Further, mastery of the task by the teacher would facilitate her use of the A-B-A method of behavior modification in which some deviant behavior is pinpointed, observation of the behavior occurs until the observer is satisfied that a stable rate of response for the behavior in question is established (Baseline), and the A-B-A approach ensues. The paradigm includes three elements: (1) the application of reinforcement contingencies to reduce or eliminate undesirable behavior and to develop more appropriate behavior, (2) a reversion to the natural social situation in which previous reinforcement contingencies are followed, and (3) the reinstatement of the new reinforcement contingencies.

The value of this A-B-A approach in aiding teachers, as well as its general nature, has been aptly summarized by Bijou and Baer (1963, p. 229): "The frequency with which these simple techniques work, and work powerfully, reinforces the belief that in fact a great deal of the complexity of real-life situations can be analyzed in simple terms. Often only a few basic principles of stimulus and response interactions are involved. It is the discovery of the precise nature of the stimuli and responses which are the details of these simple principles that is the key problem to analysis."

It is the contention of the investigator that the process of verbal conditioning, such as embodied in the present learning task, introduces the teacher to the functional analysis of behavior, suggests its powerful nature, contributes to the analysis itself, and serves as a rich clinical tool in its own right for the diagnosis, prognosis, and treatment of behavior.

Reinforcement history. The goal of operant conditioning is complete control over environmental contingencies to the extent that only the treatment effects are varied. In field experimental approaches, attainment of control at this level is highly improbable. In the absence of this control, a behavioral scientists seeks out the antecedent conditions lending to predictable behavior. Such a search lends to the study of the individual's reinforcement history. As documented before, some research workers regard the lack of control over this variable as the major contributor to the confusion so prevalent among the findings of verbal conditioning studies. Others (Patterson, 1965) regard knowledge about reinforcement history as a helpful base of operation for aiding parents in the development of reinforcement procedures to shape their children's behavior. The significance of data from reinforcement histories should be apparent.

The search for these data logically leads to an analysis of the methods applicable for such purpose. Three useful procedures are interviews, questionnaires, and direct observations. Precautions and values for each methodology are as follows.

The parental interview suffers from two major deficiencies. As Patterson (1965, p. 271-272) has asserted, the research clearly shows that parental child-rearing practices change as a function of the child's age and that parents do not accurately recall earlier child-rearing practices. Consequently, assessment of parents' current practices cannot be used as an index of their earlier handling of the child. "Because of these restrictions on interview data it can only be assumed that the assessment of current parental practices clarifies the nature of parent-child interactions which maintain the child's responsiveness to social stimuli."

A second major source for procuring data on reinforcement histories is the questionnaire.[1] Questionnaires of child-rearing practices suffer from the same limitations as the parent interview but present an additional difficulty in that the parents may answer questions in terms of the most socially desirable response.[2] The social desirability factor looms large in questionnaire data.

---

[1] A forthcoming publication contains descriptions of a large number of unpublished structured interviews, questionnaires, and objective frames of references for observation: Johnson, Orval G., and Bommarito, James W., Child Measurement Manual, San Francisco, Calif.: Jossey-Bass (in preparation).

[2] Some questionnaires require the child to answer in terms of his perceptions of parental practices. Such data furnish the possibility of an interesting test of phenomenological psychology which emphasizes the importance of the child's vantage point (e.g., Dreikurs, 1957). The question arises, then, as to whether the child's perceptions or direct data from the parents would be more closely related with an objective external criterion of behavior, such as conditionality or a teachers rating scale. To the investigator's knowledge, this test has not been made in the research literature.

Preliminary findings based on these data can be used to identify salient areas of interaction, which can then be studied more intensively, using direct observation techniques within which the A-B-A approach would be imbedded. A structured task by which the parent dispenses the reinforcement to the child may be a means of facilitating the procurement of relevant data in the observational period (Patterson, 1965). A conditioning task fits the above purpose quite well. In the investigator's estimation, a conditioning task would also provide an objective measure of progress by the parent in her handling of the child. In other words, if she is applying reinforcement techniques in an effective manner, her value as a reinforcer to the child should be enhanced. This enhancement should be reflected in the conditionality of the child by the parent in a structured learning task. Thus, a conditioning task can facilitate an analysis of behavior and serve as an index of progress in parent-child relations.

In conclusion, the crucial nature of knowledge concerning a child's reinforcement history for the major hypothesis of these studies should be mentioned. As final comments, two very potentially fruitful lines of research relating to reinforcement histories will also be discussed.

The major hypothesis of these studies has inferred that maladjusted children had reinforcement histories inferior to those of adjusted pupils. But until the relationship between reinforcement histories of the children and their gain scores is studied, the psychological meaning of this experimental behavior can hardly be interpreted with any appreciable degree of certainty.

In a closely associated area, gain scores could be related to significant age interactions. As Becker has pointed out "...age interactions are particularly interesting because they offer further opportunities to test hypotheses about different reinforcement histories. For example, can it be demonstrated that termination of negative reinforcement is used more frequently by parents of older children to influence their behavior than by parents of younger children" (1962, p. 3)? Interesting variations of the above suggestion may also be attempted. In this connection, Becker's (1962, p. 4) recommendations are given in full because of their significance. "Future research might well study the social reinforcement of the child's own parents under controlled experimental conditions. Variations of this approach might study the effect of a single parent on a sample of children, or compare the reinforcing value of a parent with that of trained adults who could serve as a standard. Another approach would be to make inferences about the reinforcement value of parents for various activities from a knowledge of their child rearing practices and to test these inferences under controlled experimental situations."

Bandura and Walters (1963, p. 12) have also called attention to the principle that interactions between reinforcement contingencies and age levels should vary in efficiency for the simple reason that behaviors which are considered desirable and thus reinforced in our culture vary with different age levels. In any case regardless of the emphases followed in studying age interactions, the investigator contends that the simple discrimination learning task and the learning criterion utilized in these studies appear to be quite useful for Becker's suggestions on controlled experimental situations.

In closing, the interconnections between social responsiveness of children to the parent of the same or opposite sex and their reinforcement histories offer a wealth of research possibilities and clinical data. In this context, Patterson (1965) has presented abundant evidence that responsiveness to social approval of the opposite-sexed parent is related to poor adjustment behavior. However, responsiveness to the sex of the parent, as determined by parental interviews, has indicated also that it depends on the child-rearing practices. "...permissive parental practices are associated with greater responsiveness to social reinforcers delivered by the same-sexed parent. Restrictive parental practices are associated with greater responsiveness to social approval of the opposite-sexed parent and to social disapproval dispensed by either parent" (p. 175). This total picture again suggests that inferences can be made about the reinforcement value of parents for controlled conditioning tasks from a knowledge of their child-rearing practices. The implications of the data from such tasks, particularly if they exhibit predictive and/or concurrent validity, could be plentiful for the behavioral management of children.

Schedules of reinforcement. As they pertain to these studies, the question of schedules arises because of the possibility that the inclusion of intermittent reinforcement within a conditioning task might improve its effectiveness. Earlier in the research report, data were cited which supported the preference of continuous reinforcement in conditioning tasks with young children. The investigator finds it difficult to accept this premise for some very good reasons. First, some clear-cut research evidence exists which indicates the effectiveness of intermittent reinforcement. The study by Kelly (1965), for example, disclosed that negative reinforcement on a fixed interval schedule was a very effective process for shaping the behavior of kindergarten children. Secondly, partial reinforcement more closely resembles real-life situations than negative reinforcement. As one illustration, Holland and Skinner (1961, pp. 133-134) have stressed that affection, attention, and approval furnish such subtle stimuli that they often go unnoticed by the recipient. Accordingly, behavior reinforced by these activities tends to be intermittently reinforced. As another example, Bandura and Walters (1963, p. 7) have documented in considerable detail the manner in which the reinforcements of daily living occur on an intermittent basis. As they further emphasized, intermittent reinforcements do not predominate solely due to the unreliability of human beings. They also predominate because of the complexity of social demands. Finally, there is abundant data on resistance to extinction following intermittent reinforcement (Estes, 1962, p. 115; McCandless, 1961, p. 141.).

The significance of these data should be clear. If intermittent reinforcement more closely simulates real life than continuous reinforcement, it should follow that it would serve as a better index of adjustment status as defined in this study in that the concept of stimulus generalization would apply more appropriately in this case. Further, in view of the resistance of behavior to extinction under conditions of partial reinforcement, it would seem logical to conclude that the effectiveness of a screening instrument for emotional disturbance which rests on susceptibility to conditioning would prove to be more reliable under this type of reinforcement.

In short, intermittent reinforcement conceivably could have two distinct advantages over continuous reinforcement. One, susceptibility to reinforcement to adults would be more apparent in that the experimental situation would have more common elements with the reinforcement contingencies in the real life of the child. Two, the influence of negative reinforcement by an adult on the responses of children possibly would have greater duration and thus prove to be more reliable. Accordingly, the logic of using intermittent reinforcement is sufficiently compelling to suggest that, at some future time, the learning task should incorporate this approach in order to compare its effectiveness with continuous reinforcement.

VII. CONCLUDING COMMENTS

The studies under consideration, through a process of concurrent validation, attempted to develop the initial stages of a practical, compact, and efficient screening instrument. The major thesis of these investigations was that the gain scores on the learning task criterion, obtained under conditions of mild verbal punishment, would differentiate the adjusted children from the maladjusted ones. In this case, the learning task criterion was the number of human responses. The design of this replication experiment was as follows: from a total kindergarten population of more than 700 children, 224 pupils were randomly selected, with 112 pupils in each sex category. A modified stratum was also used for teachers so that all in the school system would become equally involved. Their task was to rate the randomly selected children on three personality scales. The nature, rationale, and value of these have been described fully. In general, they included items encompassing the prevalently used research definitions of emotional disturbance, as well as the common problems for which referrals are made to the mental health specialists within the schools.

In the rating procedure itself, each child was rated two to three weeks apart so as to minimize the effect of previous ratings on later ones. The children who scored at the P_{73} level or greater or its equivalent on all three scales were designated as maladjusted; those who rated at the P_{23} level or lower or its equivalent on all three scales were classified as adjusted. Subsequently, the children within each sex adjustment category were randomly assigned to each examiner, with randomization also provided for alternates. Each examiner tested 15 children of each sex group within both adjustment categories for a total of 30 children. The testing procedure for each examiner included several conditions. First, adherence to good experimental conditions was followed. These have been elaborated previously. Secondly, the examiners administered the Peabody Picture Vocabulary Test first in the interest of motivating the child to task participation. Thirdly, the examiners then gave the simple discrimination learning task. This consisted of a series of 25 cards with a picture of a toy and human on each card. The presentation of all 25 cards constituted a block. Each child received four blocks of stimulus cards. On the first block, the examiner simply asked the subject to give his preference for a toy or a human figure. The base score was then the number of humans selected on block one. Beginning with the second block and continuing through the fourth block, the examiner said "that's bad" every time the S selected a toy. In the actual scoring of human preferences, the base score was subtracted from succeeding block scores and a constant of 25 then added. This process was followed to eliminate possible initial variability. Analyses of variance and $\underline{t}$ tests were based on these "gain scores" for Blocks 2, 3, and 4. In the original study, the data revealed that there was a statistically significant difference in gain scores between adjusted and maladjusted groups. In the replication study, these significant differences emerged only after statistical adjustments were made for possible procedural errors. Other major related findings of some significance included the following:

1. In both studies, the maximization of the effects of negative reinforcement occurred over three blocks. It appeared that negative reinforcement by mild verbal punishment on a schedule of continuous reinforcement had a decided effect on children's behavior at the kindergarten level. In the first study, failure of a child to show gain scores was quite rare. In the replication study, this failure was much more common but marked by perseverative behavioral tendencies when it did occur. Almost all children exhibiting this behavior had been rated as maladjusted. In either study, the simple learning task offered numerous opportunities for clinical data.

2. No sex differences emerged. Examiner differences did appear, however, in main effects. In the original research, these were very large and rested principally with the maladjusted group, particularly the male-maladjusted. In the replication, main effects related to scores of culturally deprived children, regardless of adjustment status. Lowered gain scores in both instances were related to the examiners with higher levels of training and/or experience. Finally, interaction effects between examiners and adjustment status of the children were quite insignificant.

3. The relationship between mental age and gain scores for the entire sample was found to be quite negligible. Since several studies have obtained this relationship, future research with the learning task in this study could justifiably ignore this variable. Nevertheless, the variable of intelligence remains an important one in verbal conditioning, particularly if task complexity is an issue.

4. The 1964 study gave some support to Bower's statistics on a relationship between the type of paternal occupation and the adjustment of the child. The replication did not cross-validate this relationship but did find that adjustment status, as measured by the teacher rating scales, was highly associated with scores in socioeconomic status. Methodological considerations and theoretical premises were presented as explanations for the inconsistency of the data. An alternative consideration offered, which was quite preferable to the investigator, was that variables other than the relationship between social class membership and the criterion measure could conceivably be of greater theoretical and practical significance. These were outlined in some detail. The deprivation index especially appeared to be quite relevant for a functional analysis of behavior.

In discussing the findings, considerable attention was devoted to the significance of gain scores under negative reinforcement. These related to three principal areas: (1) to the bases not only for the effectiveness of verbal punishment in general but also to the form it assumed in this study in particular, (2) to the use of punishment in the socialization process, and (3) to the possible reasons for the discriminatory power of negative reinforcement of adjustment status.

With respect to this last point, the possibility that both reward and punishment are generalized reinforcers was especially important. An additional area of discussion related to possible reasons for the rampant confusion in the literature of verbal conditioning, primarily to provide guidelines in future research. Though numerous variables contributed to these inconsistencies, task structure, the reinforcement histories of subjects, and examiner differences appeared to be especially important.

The last sections of the discussion presented suggestions for further research. These centered about immediate research needs to establish a practical screening instrument and numerous possibilities for extended applications. In line with suggestions by prominent research workers, these extended applications were subsumed under the major variables of (a) subject, (b) experimenter, and (c) situational, social, and contextual cues.

In conclusion, the simple discrimination learning was able to differentiate the criterion groups on the basis of susceptibility to mild verbal punishment. The prediction of adjustment status was rather definitive in the earlier research but rested on ad hoc statistical calculations in the replication with adjustments for procedural error. In effect, these studies provided construct validity through inference to the assumption that socialization is a function of reinforcement histories. Complete validity, however, cannot be established until the direct relationship between reinforcement histories and susceptibility to negative is confirmed. Suffice it to say, at this point, that construct validity for the learning task as a screening instrument for emotional disturbance among kindergarten children received support from at least two sources of data: (1) its relationship to practical criteria (concurrent validity), and (2) the emergence of group differences.

In addition to the support of the principal assumption of this paper by inferences from the indices of construct validity, several other significant matters deserving mention related to practical and/or theoretical consideration on the value of this study in general and the screening instrument in particular.

1. As Anderson and Harris have emphasized, a good, compact, screening instrument for emotional disturbance has yet to be developed. In a related area, Salzinger (1959, p. 88) has stated, from the standpoint of a learning theorist, as follows: "Investigations on change in the susceptibility to reinforcement due ... to differences in population have yet to be made." It is hoped that these studies have provided a beginning at least for the development of an instrument for emotional disturbance that will meet the needs in both of the areas above with respect to pragmatic and/or theoretical applications.

2. It cannot be emphasized too strongly that the investigator would abhor the use of conditioning tasks simply to select children who are "disturbed". The chief value of the data from such tasks is to sharpen intervention techniques in shaping more appropriate behavior among children with problems in social adaptation. On the basis of the evidence from the studies presented here and the research literature as a whole, it appears that verbal conditioning is useful for diagnostic, therapeutic, and prognostic purpose. Indeed, the contention by some workers who have done extensive research in the area (e.g., Krasner) is that verbal conditioning is a treatment in its own right and prototypical of social situations. Whether one finds these conclusions acceptable or not, this investigator believes that the enthusiasm of individuals for the benefits of verbal conditioning rests on solid ground. It appears to be a tool that will provide sorely needed clinical help to teachers, mental health workers, parents and to children themselves. The testability and operationalism of its concepts, moreover, provide a refreshing stimulus to the instigation of research, as exemplified by the definition of emotional disturbance in the present conditioning task as responsiveness to social stimuli in a manner that is objective, explicit, measurable, and repeatable.

As a matter of fact, probably the greatest value of this research for the investigator was in the vistas it opened up for future research. As Sarason (1965, p. 214) has affirmed, the sensitivity of verbal conditioning to such complex variables as the examiner, the subject, the social situation, and interactional effects underlies its usefulness as a research device.

3. The import of conditioning in general and of this study in particular may also be associated with the theoretical position workers in the behavioral sciences adopt. Admittedly, it is quite doubtful that a simple conditioning paradigm is going to account for all of the variance in personality traits. The neglect of cognitive factors, for example, will impose limits on the efficiency of conditioning tasks in accounting for this variance. Notwithstanding this serious deficiency, Patterson (1965, p. 176) echoed the position of many learning theorists when he stated: "It is our strong conviction that many of the personality traits and problems that have been ascribed to such determinants as 'ego', 'latent anxiety', and 'aggressive drive', can more efficiently be accounted for by a straightforward application of the conditioning paradigm ...." In a similar vein, it is the author's conviction that "emotional disturbance" as a learned pattern of behavior can also be accounted for by this paradigm.

A final issue has to do with the modern schism in psychology today between science and professionalism. On this point, the investigator's views are strictly in accord with those which have been expressed recently by Ericksen (1963). As he stated, this schism is all too prevalent today, though it really belongs in the "horse and buggy" era. The crucial issue is the utilization of science. For this purpose, validity construct is not established until we get feedback from the natural behavioral setting. Hopefully, this study has provided this feedback to some extent for the assumption that socialization is a function of reinforcement history. Herein perhaps lies its greatest contribution for the present and its promise for the future.

REFERENCES

Anderson, John E. The prediction of adjustment of time. Ira Iscoe and Harold W. Stevenson (Eds.). *Personality development in children*. Austin, Texas: University of Texas Press, 1960. Pp. 28-72.

Anderson, John E. and Harris, Dale B. *A survey of adjustment over time*. Minneapolis, Minnesota: Institute of Child Development, 1959.

Arkin, Herbert and Colton, Raymond R. *Tables for statisticians*. (2nd. ed.) New York: Barnes Noble, 1963.

Atkinson, John W. Personality dynamics. *Annual Review of Psychology*, 1960, 11, 255-290.

Ansubel, David. *Theory and practice in child development*. New York: Grune and Stratton, 1958.

Balow, Bruce. A program of preparation of teachers of disturbed children. *Exceptional Children*, 1966, 32, 455-468.

Bandura, Albert. Behavioristic psychotherapy. Palo Alto, Calif: Stanford University. Unpublished manuscript, mimeographed (undated).

Bandura, Albert. Punishment revisited. *Journal of Consulting Psychology*, 1962, 26, 298-301.

Bandura, Albert, and Walters, Richard. *Adolescent agression*. New York: Ronald Press, 1959.

Bandura, Albert, and Walters, Richard H. *Social learning and personality development*. New York: Holt, Rinehart & Winston, 1963.

Baughman, E. Earl, and Dahlstrom, W. Grant. *Negro and White children: a psychological study in the rural south*. New York: Academic Press, 1968.

Becker, Wesley. Developmental psychology. *Annual Review of Psychology*, 1962, 13, 1-34.

Beller, E. Kuno, and Turner, John. Dependency and aggression: sex differences in "normal" and "emotionally disturbed" children. Presentation at the annual convention of the American Psychological Association. St. Louis, Missouri, September, 1962.

Bergel, Ernest. *Social stratification*. New York: McGraw-Hill, 1962.

Bettleheim, Bruno, and Janowitz, Marie. *The dynamics of prejudice*. New York: Harper Publ., 1950.

Bijou, Sidney W., and Baer, Donald M. The laboratory-experimental study of child behavior. In Paul H. Mussen (Ed.), *Handbook of research methods in child development*. New York: John Wiley & Sons, 1960. Pp. 140-197.

Bijou, Sidney W. and Baer, Donald M. Some methodological contributions from a functional analysis of child development. In Lewis P. Lipsett and Charles C. Spiker (Eds.), *Advances in child development*. (Vol. I.) New York: Academic Press, 1963. Pp. 197-231.

Bommarito, James W. Conditioning by mild verbal punishment as a predictor of adjustment in kindergarten. Unpublished doctoral dissertation, Wayne State University, 1964.

Bower, Eli. *A process for the early identification of emotionally disturbed children*. Sacramento, Calif: State Dept. of Education Bulletin, 1958.

Bower, Eli M. *Early identification of emotionally handicapped children*. Springfield, Illinois: Charles C. Thomas, 1960.

Bower, Eli M., Lambert, Nadine M., Vleet, Phyllis L. Whitney, Gilbert, and Fisher, Konrad. *The education of the emotionally handicapped*. Calif: State Dept. of Education Bulletin, 1961.

Bower, Eli, and Lambert, Nadine. *A process for in-school screening of children with emotional handicaps: technical report for school administrators and teachers*. Princeton, New Jersey: Educational Testing Service, 1961.

Bower, Eli, and Lambert, Nadine. *A process for in-school screening of children with emotional handicaps: manual for school administrators and teachers*. Princeton, New Jersey: Educational Testing Service, 1961a.

Bronfenbrenner, Urie. Toward a theoretical model for the analysis of parent-child relationships in a social context. In John C. Glidewell (Ed.), *Parental attitudes and child behavior*. Springfield, Illinois: Charles C. Thomas, 1961. Pp. 90-109.

Cook, Lloyd A., and Cook, Elaine. *A sociological approach to education*. (3rd ed.) New York: McGraw-Hill, 1960.

Cooper, Saul, Ryan, William, and Hutcheson, R. R. Classroom screening for emotional disturbance. Presentation at the annual convention of the American Psychological Association, September, 1959.

Crandall, Vaughn. Some problems of personality development research. Child Development, 1956, 27, 197-203.

Hilgard, Ernest R. Theories of learning (2nd ed.) New York: Appleton-Century-Crofts, 1956.

Cronbach, Lee J. Essentials of psychological testing. (2nd. ed.) New York: Harper & Bros., 1960.

Dreikurs, Rudolph. Psychology in the classroom. New York: Harper, 1957.

Ellis, Albert, Requisite conditions for basic personality change. Journal Consulting Psychology, 1959, 23, 538-540.

Estes, W. K. Learning theory. Annual Review of Psychology, 1962, 13, 107-144.

Franks, Cyril. Conditioning and abnormal behavior. In Hans J. Eysenck (Ed.), Handbook of abnormal psychology: an experimental approach. New York: Basic Books, 1961. Pp. 457-487

Gildea, Margaret. Material attitudes and general adjustment in school children. In John C. Glidewell (Ed.), Parental attitudes and child behavior. Springfield, Illinois: Charles C. Thomas, 1961. Pp. 42-89.

Goertzen, S. M. A study of teachers' and psychologists' ability to predict seventh graders' opinion of certain behavior of their peer group. Journal of Educational Psychology, 1957, 48, 166-170.

Gram, Armin. Instructor, Merrill-Palmer Institute, Interview, Detroit, Michigan, June 7, 1962.

Hahn, Milton. Forgotten people: the normal individual and, and in, professional psychology. American Psychologist, 1962, 47, 700-705.

Hoffman, M. L. Power assertion in the parent and its impact on the child. Child Development, 1960, 31, 129-143.

Holland, James G. and Skinner, B. F. The analysis of behavior. New York: McGraw-Hill, 1961.

Hollister, William G. and Goldston, Stephen E. Classes for the emotionally disturbed: considerations for planning. Washington D. C.: Council for Exceptional Children, 1962.

Jenness, Arthur. Personality dynamics. Annual Review of Psychology, 1962, 13, 479-514.

Kelly, Richard. Comparison of the effects of positive and negative reinforcement in an operant learing task. *Journal of Educational Psychology*, 1966, 57, 307-310.

Kennedy, Wallace A. and Willcutt, Herman C. Praise and blame as incentives. *Psychological Bulletin*, 1964, 62, 323-332.

Kinget, C. Marion. *The drawing-completion test*. New York: Grune and Stratton, 1952.

Krasner, Leonard. Verbal Contitioning and psychotherapy. In Leonard Krasner and Leonard P. Ullmann (Eds.), *Research in behavior modification: new developments and implications*. New York: Holt, Rinehart, & Winston, 1965. Pp. 211-228.

Long, Peter J. Conditioning and learning theory psychotherapy. Presentation at the annual convention of the American Psychological Association, St. Louis, Missouri, September, 1962.

Long, Nicholas, Morse, William C., and Newman, Ruth G. (Eds.), *Conflict in the classroom*. Belmont, Calif.: Wadsworth, 1965.

National Opinions Research Center. Jobs and occupations: a popular evaluation. In Reinhard Bendix and Seymour M. Lipset (Eds.), *Class, status, and power*. Glencoe, Illinois: The Free Press, 1953. Pp. 411-426.

Marshall, Hermine H. The effect of punishment on children. *Journal of Genetic Psychology*, 1965, 106, 23-33.

McCandless, Boyd. *Children and adolescents*. New York: Holt, Rinehart, and Winston, 1961.

Maccoby, Eleanor R. The choice of variables in the study of socialization. *Sociometry*, 1961, 24, 357-371.

Matarazzo, Joseph D. Psychotherapeutic processes. *Annual Review of Psychology*, 1956, 16, 181-224.

McMabola, Marie. Research and mental health in the schools. *Social Work*, 1961, 6, 82-86.

Michael, Jack and Meyerson, Lee. A behavioral approach to guidance and counseling. In Ralph Mosher, Richard F. Carle, and Chris D. Kehas (Eds.), *Guidance: an examination*. New York: Harcourt, Brace, and World, 1965. Pp. 24-48.

Milholland, John E. Theory and teaching of assessment, *Annual Review of Psychology*, 1964, 15, 311-346.

Mussen, Paul. Developmental psychology. Annual Review of Psychology, 1960, 11, 439-478.

Mussen, Paul, and Conger, John L. Child development and personality. New York: Harper Brothers, 1956.

Newman, Ruth G., Redl, Fritz, and Kitchener, Howard L. Technical assistance in a public school. Washington, D. C.: Washington School of Psychiatry, 1961.

Patterson, Gerald. Responsiveness to social stimuli. In Leonard Krasner and Leonard P. Ullmann (Eds.), Research in behavior modification: new developments and implications. New York: Holt, Rinehart, and Winston, 1965. Pp. 157-178.

Peterson, Donald R. Behavior problems of middle childhood. Journal of Consulting Psychology, 1961, 32, 151-162.

Popham, W. James. Educational statistics: use and interpretation. Evanston, Illinois: Harper & Row, 1967.

Quay, Herbert, C., Morse, William C., and Cutler, Richard L. Personality patterns of pupils in special classes for the emotionally disturbed. Exceptional Children, 1966, 36, 215-220.

Quay, H. C., Sprague, R. T., Shulman, H. S., and Miller, A. T. Some correlates of personality disorder in a child guidance clinic sample. Psychology in the Schools, 1966, III, 44-47.

Quay, Herbert C., and Quay Lorence C. Behavior problems in early adolescence. Child Development, 1965, 36, 215-220.

Radke, Marie J. The relation of parental authority to children's behavior and attitudes. Minneapolis, Minnesota: Univ. of Minn. Press, 1946.

Reiss, Albert J., Jr. Occupations and social status. New York: The Free Press, 1961.

Reiter, Richard. Director of Computer Services, Illinois State University, Interview, Normal, Illinois, October 10, 1968.

Rotter, Julian B. Social learning and clinical psychology. Englewood Cliffs, N. J.: Prentice-Hall, 1954.

Rubin, Eli. Personal communication, November 15, 1966.

Rubin, Eli M., Simson, Clyde B., and Betwee, Marcus C. Emotionally handicapped children and the elementary school. Detroit, Michigan: Wayne State University Press, 1966.

Salzinger, Kurt. Experimental manipulation of verbal behavior: A review. *Journal of General Psychology*, 1959, 61, 65-94.

Salzinger, Kurt, and Portnoy, S. Verbal conditioning in interviews: application to chronic schizophrenics and relationship to prognosis for acute schizophrenics. *Journal of Psychiatric Research*, 1964, 2, 1-9.

Sarason, Irvin G. The human reinforcer in verbal behavior research. In Leonard Krasner and Leonard P. Ullmann (Eds.), *Research in behavior modification: new developments and implications*. New York: Holt, Rinehart, and Winston, 1965. Pp. 229-243.

Scott, William A. Research definitions of mental health and mental illness. *Psychological Bulletin*, 1958, 55, 1-45.

Scott, William. Social psychological correlates of mental illness and mental health. *Psychological Bulletin*, 1958a, 55, 65-85.

Sears, Robert R., Maccoby, Eleanor E., and Levin, Harold. *Patterns of child rearing*. Evanston, Illinois: Row Peterson, 1957.

Siegel, Sidney. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

Smith, Milton. *A simplified guide to statistics*. (3rd ed.) New York: Holt, Rinehart, and Winston, 1962.

Spence, Janet T., and Dunton, Moira C. The influence of verbal and nonverbal reinforcement combinations in the learning of middle- and lower-class preschool children. *Child Development*, 1967, 38, 1177-1186.

Spiker, Charles C., and McCandless, Boyd R. The concept of intelligence and the philosophy of science. *Psychological review*, 1954, 61, 255-266.

Sullivan, Paul. The effect of verbal reward and punishment on concept elicitation in children. Unpublished doctoral dissertation, Wayne State University, 1960.

Swanson, Guy E. Determinants of the individual's defense against inner conflict. In John C. Glidewell (Ed.), *Parental attitudes and child behavior*. Springfield, Illinois: Charles C. Thomas, 1961. Pp. 5-41.

Terrell, G., Durken, W., and Wiesley, M. Social class and the nature of the incentive in discrimination learning. *Journal of Abnormal and Social Psychology*, 1959, 59, 270-272.

Thompson, George S. Developmental psychology. Annual Review of Psychology, 1959, 10, 1-42.

Ullmann, Leonard P., and Krasner, Leonard P. Introduction. In Leonard P. Ullmann (Eds.), Case studies in behavior modification. New York: Holt, Rinehart, and Winston, 1965. Pp. 1-67.

U. S. Bureau of the Census. Methodology and scores of socioeconomic status. Working paper No. 15. Washington, D.C., 1963.

Warburton, F. W. Educational psychology. Annual Review of Psychology, 1962, 13, 371-414.

White, Mary A. and Harris, Myron M. The school psychologist. New York: Harper and Brothers, 1961.

Wirt, Robert D. and Wirt, Anne L. Psychotherapeutic processes. Annual Review of Psychology, 1963, 14, 365-390.

Wright, M. Erik. Abnormalities of behavior. Annual Review of Psychology, 1957, 8, 269-308.

Whiteman, Martin and Deutsch, Martin. Social disadvantage as related to intellective and language development. In Martin Deutsch, Irwin Katz, and Arthur R. Jensen (Eds.), Social class, race, and psychological development. New York: Holt, Rinehart, and Winston, 1968. Pp. 86-114.

Zigler, Edward, and Kanzer, Paul. The effectiveness of two classes of verbal reinforcers on the performance of middle- and lower-class children. Journal of Personality, 1962, 30, 157-163.

Zigler, E., and De Tabry, J. Concept-switching in middle class, lower-class, and retarded children. Journal of Abnormal and Social Psychology, 1962, 65, 267-273.

APPENDIXES

APPENDIX A: PROBLEM CHECKLIST

FACE SHEET (for original selections, first rating only)

1. Name of child rated: ______________________________

2. Birthdate: ______________________ 3. Sex: ______ M ______ F

4. Occupation of father (please be as specific as possible: e.g., not skilled trade or worker but tool and die worker)

____

5. How long have you known this child: (please give total time you have known child in months): ______________________

6. Name of school: ______________________________

7. Name of principal: ______________________________

8. Name of teacher rating child: ______________________

9. Signature of teacher:

____

10. Date rating was completed and forwarded: ______________

11. Check time of day child attends school: A.M. ________ P.M. ______

INSTRUCTIONS:

1. Items 1, 2, 3, 6, 7, and 8 will be completed for all teachers in advance for originally selected children.

2. All teachers will need to complete items 4, 5, 9, 10, and 11 (if not given).

3. If child is an alternate, please use face form for alternates.

4. In the case of alternate ratings, staple face sheet for original selection to alternate form of child who has replaced him and mark conspicuously on the original the words NOT RATED.

5. If you are certain that the data completed in advance is not accurate, please feel free to make the necessary changes.

PROBLEM CHECKLIST

DIRECTIONS: Below is a list of some behavioral characteristics which research has shown to be common among many kindergarten children. Rate each item on the scale as directed. But please do not feel any concern on your part in describing a child's difficulties through these items. The results will be evaluated by a competent professional worker with considerable training and experience in psychology and education who will safeguard the confidentially of the material and use the group results strictly for research purposes.

In completing the scale, please rate only those children whom you have known for at least two months, even if it means rating an alternate.

Your care and thoroughness in the completion of this scale will be appreciated.

PROBLEM CHECKLIST

Name of Child:__________________________________Sex:__________________

Birth Date:______________________________Date of Rating:_______________

Teacher:_________________________________School:______________________

Please indicate which of the following constitute problems, as far as this child is concerned. If an item does not constitute a problem, encircle zero; if an item constitutes a mild problem, encircle the one; if an item constitutes a severe problem, encircle the two. Please complete every item.

0 1 2 1. Thumb-sucking

0 1 2 2. Restlessness, inability to sit still

0 1 2 3. Attention-seeking, "show-off" behavior

0 1 2 4. Skin Allergy

0 1 2 5. Doesn't know how to have fun; behaves like a little adult

0 1 2 6. Self-conscious; easily embarrassed

0 1 2 7. Headaches

0 1 2 8. Disruptiveness; tendency to annoy and bother others

0 1 2 9. Feelings of inferiority

0 1 2 10. Dizziness, vertigo

0 1 2 11. Boisterousness, rowdiness

0 1 2 12. Crying over minor annoyances and hurts

0 1 2 13. Preoccupation; "in a world of his own"

0 1 2 14. Shyness, bashfulness

0 1 2 15. Social withdrawal, preference for solitary activities

0 1 2 16. Dislike for school

0 1 2 17. Jealousy over attention paid other children

0 1 2 18. Difficulty in bowel control, soiling

0 1 2 19. Prefers to play with younger children

0 1 2 20. Short attention span

0 1 2 21. Lack of self-confidence

0 1 2 22. Inattentiveness to what others say

0 1 2 23. Easily flustered and confused

0 1 2 24. Lack of interest in environment, generally "bored" attitude

0 1 2 25. Fighting

0 1 2 26. Nausea, vomiting

0 1 2 27. Temper tantrums

0 1 2 28. Reticence, secretiveness

0 1 2 29. Truancy from school

0 1 2 30. Hypersensitivity; feelings easily hurt

0 1 2 31. Laziness in school and in performance of other tasks

0 1 2 32. Anxiety, chronic general fearfulness

0 1 2 33. Irresponsibility, undependability

0 1 2 34. Excessive daydreaming

0 1 2 35. Masturbation

0 1 2 36. Hay fever and/or asthma

0 1 2 37. Tension, inability to relax

0 1 2 38. Disobedience, difficulty in disciplinary control

0 1 2 39. Depression, chronic sadness

0 1 2 40. Uncooperativeness in group situations

0 1 2 41. Aloofness, social reserve

0 1 2 42. Passivity, suggestibility; easily led by others

0 1 2 43. Clumsiness, awkwardness, poor muscular coordination

0 1 2 44. Stuttering

0 1 2 45. Hyperactivity; "always on the go"

0 1 2 46. Distractibility

0 1 2 47. Destructiveness in regard to his own and/or others' property

0 1 2 48. Negativism, tendency to do the opposite of what is requested

0 1 2 49. Impertinence, sauciness

0 1 2 50. Sluggishness, lethargy

0 1 2 51. Drowsiness

0 1 2 52. Profane language, swearing, cursing

0 1 2 53. Prefers to play with older children

0 1 2 54. Nervousness, jitteriness, jumpiness; easily startled

0 1 2 55. Irritability; hot-tempered, easily aroused to anger

0 1 2 56. Enuresis, bed-wetting

0 1 2 57. Stomach-aches, abdominal pain

0 1 2 58. Specific fears, e.g., of dogs, of the dark

Please note here any problems not mentioned above.

MAJOR SUBSCALES OF PROBLEM CHECKLIST

I. Factor: Conduct Problem (Unsocialized Aggression, Psychopathy) (17 items)

Item	Average factor loading
2	.60
3	.62
8	.70
11	.66
16	.35
17	.40
25	.63
27	.51
33	.57
38	.70
40	.62
45	.51
47	.57
48	.59
49	.57
52	.48
55	.59

II. Factor: Personality Problem (Neurotic-Distrubed) (14 items)

5	.41
6	.55
9	.60
12	.33
14	.51
15	.50
21	.60
23	.53
28	.38
30	.50
32	.53
37	.34
39	.48
41	.33

(Quay and Peterson, unpublished, 1967)

APPENDIX B: BEHAVIOR CHECKLIST

BEHAVIOR CHECKLIST

Name of Child:__________________________________Sex:____________________

Birth Date:________________________________Date of Rating:_____________

Teacher:___________________________________School:_____________________

Please indicate which of the following constitute problems as far as this child is concerned. If an item does not constitute a problem, encircle zero, if an item constitutes a mild problem, encircle the one; if an item constitutes a severe problem, encircle the two. Please complete every item.

0 1 2 1. Very sensitive to criticism.

0 1 2 2. Expresses feelings of inadequacy about self.

0 1 2 3. Never makes self known to others.

0 1 2 4. Is excessively neat or finicky about work or possessions.

0 1 2 5. Overconforms to rules.

0 1 2 6. Aggressive in underhanded ways.

0 1 2 7. Seeks attention excessively.

0 1 2 8. Very short attention span.

0 1 2 9. Cannot work independently.

0 1 2 10. Shows signs of nervousness (nailbiting, crying, tics, rocking).

0 1 2 11. Overly preoccupied with sexual matters.

0 1 2 12. Daydreams.

0 1 2 13. Seems to fear being assertive even in ordinary ways (asking to go to toilet, defending self, making legitimate messes, joining in allowable noisy play).

0 1 2 14. Is receiving, or recommended, speech correction.

0 1 2 15. Poor coordination (trouble with buttoning, tying shoes, getting shoes on correct feet).

0 1 2 16. Cannot take turns: "Me first."

0 1 2 17. Lacks responsibility for self, always has excuse for shortcomings.

0 1 2 18. Resists limits or rules in group games.

0 1 2 19. Tendencies toward enuresis or soiling of clothing.

0 1 2 20. Very messy with work or belongings.

0 1 2 21. Negativistic: "I won't."

0 1 2 22. Difficulty in handling working materials (crayons, scissors, paste).

0 1 2 23. Considered an isolate in class.

0 1 2 24. Engages in much solitary play.

0 1 2 25. Displays infantile behavior (crawling, whining, clinging, sucking, chewing).

0 1 2 26. Makes odd noises.

0 1 2 27. Makes irrelevant or inappropriate remarks.

0 1 2 28. Misinterprets simple statements.

0 1 2 29. Is disoriented in space; is confused as to directions given.

0 1 2 30. Shows excessive fantasy preoccupation.

0 1 2 31. Tendencies toward primitive hostilities, temper tantrums, wild destruction.

0 1 2 32. Holds back in free play.

0 1 2 33. Antisocial tendencies (steals, lies, destroys property, bullies, defies, resents discipline).

0 1 2 34. Frequently tardy, frequently absent.

0 1 2 35. Poorly cared for before leaving for school.

0 1 2 36. Easily fatigued.

0 1 2 37. Often ill; other physical problems

0 1 2 38. Feigns illness.

0 1 2 39. In academic area, evidence of underachievement, or overachievement, in relation to ability.

0 1 2 40. Child is under professional guidance

0 1 2 41. Extremely lethargic, shows little interest in surrounding activity

0 1 2 42. Has unusual difficulty with speech

0 1 2 43. Is often ill

0 1 2 44. Delights in destructive acts

0 1 2 45. Aggressive towards adults who try to set limits to behavior

APPENDIX C: MINNESOTA SCALE WITH FACE SHEET

FACE SHEET: MINNESOTA SCALE

Name of Teacher: ____

Name of Child: ______________________________ Sex: ____ M ____ F

Time of day child attends school: ____ A.M. ____ P.M.

Signature of teacher:

____

Date rating completed and forwarded: __________________________

Check here if child is an alternate: ______

INSTITUTE OF CHILD DEVELOPMENT UNIVERSITY OF MINNESOTA

PERSONALITY PROFILE II

Name____________________ School____________________ Grade________

Instructions: On the line at the left of each item, write the number of the word or phrase that corresponds to your rating of the child.

____1. IN MY OPINION, THIS CHILD'S GENERAL ADJUSTMENT IS:

Poor	Fair	Average	Good	Excellent
1	2	3	4	5

____2. HOW REALISTIC IS HE?

Knows his own faults and good points	Fairly realistic	Somewhat realistic	Doesn't seem to know the score about himself	Completely unaware of what he is like
5	4	3	2	1

____3. HOW WELL DOES HE PERSIST AT A TASK?

Gives up very easily	Gives up when he has a little trouble	Takes quite a bit to make him give up	Sticks to a job when it is very troublesome	Won't give up in spite of everything
1	2	4	5	3

____4. HOW WELL DOES HE TAKE RESPONSIBILITY FOR WHAT HE DOES?

Takes responsibility for what he does	Seldom makes excuses	Sometimes alibis	Tries to pass the buck	Definitely blames others when he is in the wrong
5	4	3	2	1

____5. HOW ATTENTIVE IS HE IN SCHOOL?

Inattentive most of the time	Tends to be inattentive	Moderately attentive	Usually attentive	Very attentive
1	2	3	4	5

____6. HOW DEPENDENT IS HE?

Tackles problems very much on his own	Independent	Uses own skill first, then seeks help	Dependent	Seeks help at slightest difficulty
5	3	4	2	1

____7. HOW FLEXIBLE IS THIS CHILD?

Very easily led or influenced	Tends to "drift with the tide"	Takes sensible suggestions, rejects others	Slow to adapt to new ideas	Rigid inflexible
1	4	5	2	3

____8. HOW MUCH AT EASE IS HE?

Passive	Relaxed	Settles down after excitement	Nervous	Tense
2	5	4	1	3

____9. HOW WELL CAN HE "TAKE IT"?

Tough, nothing offends him	Insensitive hard to hurt	Has feelings, but controls them	Soft, sometimes can't take it	Touchy, very easily hurt
4	2	5	3	1

____10. HOW COMPLIANT IS HE?

Does opposite of what he is told	Often contrary, resistive	Agrees to sensible requests	Goes out of his way to please	Too anxious to please, apple polishes
2	1	5	4	3

APPENDIX D: SOCIALIZATION SCALE

FACE SHEET (for alternates, first rating only)

1. Name of child rated: ______________________________

2. Birthdate: ____________________ 3. Sex: ________ M ________ F

4. Occupation of father: ______________________________

5. How long have you known the child? ______________________________

6. Signature of teacher: ______________________________

7. Date rating was completed and forwarded: ______________________________

8. Check time of day the child attends school: ________ A.M. ________ P.M.

SOCIALIZATION SCALE

Directions: Below is a list of some behavioral characteristics that can be observed, at times, among many children in kindergarten. Also, the items below are those which the vast majority of the 27 kindergarten teachers in the East Detroit School District have rated as possibly important or very important indicators of adjustment. But please do not feel any concern on your part in describing a child's difficulties through these items. The results will be evaluated by a competent professional worker with considerable training and experience in psychology and education who will safeguard the confidentially of the material and use the group results strictly for research purposes.

Now check only those statements which you feel are really true of the child listed on the enclosed face sheet. Do not guess if you are not reasonably sure. If at all possible, rate only those children whom you have known for at least two months, even if it means rating an alternate.

Your care and thoroughness in the completion of this scale will be appreciated.

SOCIALIZATION SCALE

______1. Watches others play; seldom plays by himself.

______2. Plays alone most of the time.

______3. Cries without seeming provocation.

______4. Seeks attention excessively.

______5. Is able to listen to stories read by the teacher for short periods at least.

______6. Is able to accept "little" jobs in the kindergarten and performs them adequately.

______7. Is able to share materials.

______8. Finds it difficult to accept just blame for his faults.

______9. Regresses to babyish behavior in the face of difficulty.

______10. Dawdles to avoid a difficult task.

______11. Tries to make entry into group of children but fails.

______12. Insists that other children do as he wishes.

______13. Resents interest shown by other children, wants to be left alone.

______14. Poor in concentration.

______15. Gives up easily; lacks persistence.

______16. Has to be constantly urged to carry out routine activities.

______17. Not absorbed, nor self-sufficient in his activity.

______18. Child exhibits frequent unprovoked physically aggressive behavior against his peers, often lengthy, with many physical attacks and fights.

______19. Quarrels with other children excessively.

______20. Frequently uses hitting or other physical means to make social contacts with others.

SOCIALIZATION SCALE

______21. Often engages in destructive kinds of play--knocking down blocks, pulling things to pieces, and otherwise destroying or damaging toys and materials.

______22. Resents aid from adults.

______23. Almost always gives up before a task is completed.

______24. Leaves as soon as a task becomes difficult.

______25. Often leaves craftwork or clean-up jobs before they are completed.

______26. Rarely retaliates or tries to stand up for his rights when attacked.

______27. Rarely or never does what an adult asks.

______28. Often ignores adults requests.

______29. Rarely cooperative in following the rules and routines of the kindergarten.

______30. Does not usually comply with adult requests without extreme coercion.

______31. Child shows no interest in school work. Hates it. Openly hostile and resentful.

______32. Child has many upsets. "Hair trigger" emotions.

______33. Practically always restless. Extremely fidgety. Can't sit still even when he tries.

______34. Feelings easily hurt. Extremely self-conscious.

______35. Seems entirely blind to social approval or disapproval.

______36. Child is so impulsive that he never stops to ponder consequences, off at the drop of a hat.

______37. Content with hopelessly careless work.

______38. Overcautious, not venturesome, afraid to attempt the untried.

______39. Repressed, restrained.

SOCIALIZATION SCALE

______40. Lacks self-confidence.

______41. Helpless unless someone organizes activity for him.

______42. Wanders about aimlessly in classroom.

______43. Merely copies other children's reactions in class activities.

______44. Very critical of other children.

______45. Does not respond to friendly advances.

COMMENTS: If you care to do so, describe this child in your own terms and/or make any other general comments you feel are pertinent in the space below:

APPENDIX E: SCHOOL OBSERVATION SCHEDULE (COOPER SCALE)

ITEMS FOR SCHOOL OBSERVATION SCHEDULE

Directions: Check only those statements which you feel are really true of the child. Do not guess if you are not reasonably sure.

_______1. Irratable, touchy
_______2. Destructive
_______3. Negativistic; refuses to comply
_______4. Sullen; passively hostile
_______5. Acts noisy, silly
_______6. Bold, "fresh," talks back
_______7. Demanding; attention-seeking
_______8. Sucks thumb
_______9. Poor drawing
_______10. General Lack of coordination
_______11. Has to be shown what to do far more than average kid
_______12. Hyperactive restless
_______13. Needs help in dressing self
_______14. Needs help in going to toilet
_______15. Cannot conform to toilet schedule
_______16. General poor grooming
_______17. Shy or timid with others
_______18. Fears to participate
_______19. Left out by other children
_______20. Lacking in spontaneity and creativity
_______21. Overly conforming
_______22. Lazy, unwilling to exert self

ITEMS FOR SCHOOL OBSERVATION SCHEDULE

_______23. Inattentive

_______24. Needs much prodding and guidance

_______25. Oversensitive to criticism

_______26. A tense child

_______27. Specific fears (of being hurt, dogs, heights, etc.)

_______28. Cries in class

_______29. Insecure; shamefaced

_______30. Jealous of other children

_______31. Hard to reach and relate; tends to withdraw

_______32. Has nervous mannerisms

_______33. Doesn't understand directions

_______34. Doesn't seem to have common sense

_______35. Marked difficulty in understanding complex game rules

_______36. Marked lack of reading readiness

_______37. Marked lack of number concepts

_______38. Child is under professional guidance

_______39. Extremely lethargic, shows little interest in surrounding activity

_______40. Rarely communicates verbally

_______41. Has unusual difficulty with speech

_______42. Is often ill

_______43. Is impatient with self, children, and adults

_______44. Delights in destructive acts

_______45. Aggressive towards adults who try to set limits to behavior

APPENDIX F: STEP BY STEP REPORT ON THE PROGRESS OF THE RESEARCH

1. As the very first step, the writer contacted the superintendent of schools in Bloomington on the possibility of having the research conducted within the school system. At a series of meetings with him and two of his assistant superintendents, the investigator explained the nature of the research, its value, and its requirements for teachers. Subsequent to mutually satisfactory understandings for both parties, school officials arranged an open meeting for the investigator with teachers and principals. Argument for participation by the teachers was facilitated by a small payment for their work, a condition which was not present in the original study.

2. At this meeting, the teachers had the opportunity to ask any questions they wished. Whenever possible, these were answered as fully and frankly as possible without divulging the complete nature of the research, since such a disclosure easily could have contaminated the results. After all teachers agreed to participate, each one received a packet containing the following materials.

a. General directions for kindergarten teachers (Appendix J).

b. The appropriate number of Problem Checklists (14 for each teacher).

c. A complete roster of original and alternate ratees.

3. Approximately four months were allowed for the completion of all scales by the teachers. Each teacher then received only one scale at a time with a two to three week interval permitted between ratings so that prior judgments would not unduly influence the raters on subsequent personality appraisals. Careful tabulations were made on a master work sheet of all ratings received and missing. Upon receipt of all three sets of completed ratings, statistical calculations were applied to the data to determine the children who scored P_{73} or more or P_{27} or less or their equivalence on all three scales. These were respectively designated as maladjusted and adjusted. In the original study (Bommarito, 1964), the Socialization Scale (Appendix D) and the Cooper Scale (Appendix E) did not discriminate as well among the adjusted as they did for the maladjusted, based on extreme percentile scores of raw scores. Accordingly, in the replication study, new scales were substituted, with the Problem Checklist and the Behavior Checklist respectively replacing the Socialization Scale and the Cooper Scale. As a further precaution against the lack of discrimination described above, raw scores were transformed to normalized standard scores in a manner described by Popham (1967, pp. 36-37). Moreover, it appeared that the new scales contained a number of advantages over the former instruments.

a. The Socialization Scale had only content validity for statistical support, whereas the Cooper Scale had concurrent validity and satisfactory reliability. Both scales, moreover, had rather small samples on which they based their results (Bommarito, 1964). In contrast, the Problem Checklist had large normative groups for kindergarten children, predictive and concurrent validity, and internal consistency (Peterson, 1961). With the exception of predictive validity and test-retest reliability, the same advantages were attributable to the Behavior Checklist (Rubin et. al., 1966).

b. Both the Problem Checklist and Behavior Checklist had subscales on withdrawn and anti-social behavior, thus facilitating a check on the relationship between gain scores by conditioning to possible oversocialization and undersocialization respectively.

4. Children scoring at these extreme percentile points were randomly assigned to each examiner. The randomization of assignments for each examiner occurred within each adjustment status and sex classification (male-adjusted, male-maladjusted, female-adjusted, and female-maladjusted). Each psychologist received a typed list with a complete roster of original examinees and alternates for each of the four adjustment classification of any of the children to be examined individually by them.

5. Prior to the individual examination of the 30 children by each psychologist, the assistant superintendent sent a letter to all elementary principals and kindergarten teachers indicating his approval of the research project and making some suggestions for good testing conditions as suggested by the investigator (Appendix H). A complete testing schedule for each school, indicating the date, time of testing, the examinees to be tested, and the name of the examiner was also enclosed with this letter. At approximately the same time, a letter of explanation was forwarded to the parents of children who were examinees for the individual testing (Appendix I).

6. Two meetings were held with the two school psychologists who were to conduct the individual testing. At these conferences, each psychologist received a packet of materials containing (1) a comprehensive directive on proper testing procedure and sequence of testing; (2) a complete schedule of testing, including school, time of day, date of testing, and children to be examined within these three variables; (3) a roster of all teachers and principals; (4) a complete list of ratees, including original children to be tested and alternates; and (5) scoring data sheets for the Peabody and the learning task. Discussions then centered around any questions the psychologists had about any concern whatsoever. Emphasis was placed on the need for uniformity of administering directions, sequence of testing, and adherence to good research conditions. Information on the value, scope, purpose, and implications of the study was also revealed to the psychologists to increase motivation for participation in the research project. The writer also stressed to the psychologists that he was available at all times if any problems arose.

7. When the psychologists completed the individual testing, the statistical calculations reported in this paper were made.

8. Finally, the only step remaining will be the report of the present research to the Bloomington School System. Since the investigator has promised to make this report, it will definitely be forthcoming.

APPENDIX G: FORMS FOR RECORDING DATA FROM LEARNING TASK AND PEABODY PICTURE VOCABULARY TEST

RAW DATA SCORING SHEET FOR LEARNING TASK AND PEABODY TEST

Background Information and Peabody Test Date

Name of S: ____________________ BD:* ____________ Sex ____ M ____ F

School: ____________________ Teacher ____________________

Time of Day: ______ a.m. ______ p.m. Date Tested: ____________________

Score on Peabody: ____________ Raw Score ____________ CA ____________ MA ____________ IQ

Form of Peabody: ____

Check here if alternate has been tested: ____________

Name of Alternate: ____________________ School ____________

Name of Alternate's Teacher: ______________________________

Time of Day Alternate Attends School ________ a.m. ________ p.m.

____
Signature or Initials of Examiner

*Refers to birthdate of the child.

LEARNING TASK DATA

	BLOCK I		BLOCK II		BLOCK III		BLOCK IV	
	H*	T**	H*	T**	H*	T**	H*	T**
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								
23								
24								
25								
Totals***								

* Place check under H if human figure was chosen.
** Place check under T if toy was chosen.
*** Refers to total for number of times toys and humans were chosen for each block.

APPENDIX H: APPROVAL LETTER TO SCHOOL PERSONNEL

TO: All Elementary Principals and Kindergarten Teachers

FROM: Don Henderson

SUBJECT: Approved Research Project by James W. Bommarito

Mr. Bommarito has informed me that his research project is now ready for its final phase. This step will consist of individual examinations of children. There will be about 60 children involved for all the schools. The children selected will be taken from the classrooms individually. The two psychologists, Mr. Pat Kroll and Mr. Rudolph Breitmeyer, will administer the tests. These examination sessions will last approximately one-half hour with each child and will include two simple tests, one of which is a simple discrimination learning task. The testing schedule will last approximately two weeks and extend from May 7 through May 14. If necessary, a few additional tests may be given on May 15. A complete schedule of visitations will be forthcoming shortly.

Advance preparation of the children is neither necessary or desirable. The only advance preparation that would be necessary would be working space for the psychologists and the assignment of a reliable sixth grade girl who could act as a runner and bring the children to the testing room. If any of your children are chosen, however, would you please talk to them for a minute before they leave the room to assure them that Mr. __________ is here to spend some time with us today, that it is all right for them to go with the runner to him, and that you are pleased that they have been chosen. We might inform the children that there are games to be played.

I would discourage any questioning after they have returned to the room for it would be hard to understand since the children haven't even a vague idea of the purpose of this study.

I know that you will continue to give excellent cooperation with this project. We will be interested in the results of the study.

APPENDIX I: LETTER FROM SCHOOL OFFICIALS TO PARENTS

TO: Parents of children involved in research project

FROM: Don Henderson, Assistant Superintendent of Schools

SUBJECT: Individual Testing

With our approval, Dr. James W. Bommarito from the Special Education Department at Illinois State University is now doing a study in our schools which checks the manner in which kindergarten children learn. As part of this study, about 60 children will take two simple tests. In one scale, the children will be checked to see how many common objects they know. In the other, they will be presented a group of cards with two common objects. In each case, they will be asked to give a preference or to state their choices. Both tasks will be given by a trained adult who will read the directions individually with each child. The entire examination should take about 30 minutes.

But the fact that your child was chosen for this study should not concern you for several reasons. First, all the kindergarten children in Bloomington had a chance of being chosen. The fact that your child was chosen was just chance. Secondly, the interest of the study is in group trends or on how the children perform as a group, not as individuals. Finally, the trained workers dealing with the children will keep the results in the strictest confidence.

If you have any further questions, please feel free to call Dr. Bommarito in the Department of Special Education at Illinois State University.

APPENDIX J: GENERAL DIRECTIONS FOR KINDERGARTEN TEACHERS

GENERAL DIRECTIONS FOR KINDERGARTEN TEACHERS

A few general requests and directions are given below. It is hoped that they will promote efficiency, avoid possible misunderstandings, and assure uniform practices.

1. Be certain that you read the directions on each scale carefully before you rate the child.

2. Also, "double check" to see that the face sheets are complete insofar as this is possible and that you have signed your name.

3. Please return all items to me, at Illinois State University.

4. Preferably, it would be desirable to have each set of completed ratings back within a few days after the first day you have received them. If you find this task too formidable, please return the ratings within a week at least. Since the school year is "fast running out" and considerable individual testing remains to be done with the children, I'm sure that you can readily understand the need for this dispatch.

5. At the same time, however, please use considerable care in the completion of the scales within these reasonable time limits. It was to promote such care that the number of children to be rated and the items to be checked have been held to a minimum.

6. If an alternate is necessary for the first rating, the same individual should be rated in subsequent ratings. It is very important that the three ratings be done for the same child. Otherwise, an analysis of group trends will be virtually impossible.

7. If an alternate moves or otherwise will not be available anymore at the time of the second or third ratings, substitute another alternate from the available list with which you will be provided. Be certain that you select a substitute that you can depend on to be available.

8. In case of "gaps" in this consistency of three ratings on the same individual, please notify me on the type of scale you need. The writer himself will also attempt to check gaps through a master list which he will develop.

9. You should understand the sequence and order of the ratings. The first scale which you will be requested to complete will be the Problem Checklist; the second, the Minnesota Scale; and the third, the Behavior Checklist in that order. You will receive one set of ratings at a time. As soon as all the teachers have returned the results, you will be requested to complete the next set, etc.

APPENDIX K: DETAILS ON MAXIMIZATION OF BLOCK EFFECTS

Popham has suggested that a simple test for homogeneity of variance can be made by dividing the largest estimate of variance by the smallest value, with degrees of freedom in each group equal to N-1 (1967, p. 145). In this case, the largest variance is for Block IV, which is equal to 54.105; the smallest estimate is the variance for Block II (38.333). The F ratio, then, is 54.105/38.333 or 1.411, with df in each group equal to 59. From a table of F, a value of 1.50 or more would be required for significance at the 10 per cent level, with df of 59 in each group. Hence, the observed of 1.411 is not significant and the conclusion of homogeneity in variance can be made.

With homogeneity of variance assured, Guilford has stated that one may apply the within groups variance as an estimate of the population variance to apply to all possible pairs of means for gain scores between blocks. According to Guilford (1956, p. 263), the residual variance is the variance within sets. In the present investigation, it is 13.28 with df equal to 118 (see Table II). For 118 df, a t value of 1.671 is required for the .05 level of significance and one of 2.358 for the .01 level (Dixon and Massey, 1957, p. 384). These values are for one-tailed tests. This procedure provides a more stable estimate of the population variance and only one standard error (SE) of a difference to compute. The SE of a difference between groups is given by Guilford (1956, p. 264) as σ dm $\sqrt{\frac{2Vw.}{n}}$ Since n is 60 and Vw is 13.28, σ dm is $\sqrt{\frac{2(13.28)}{60}}$ or .4423. Consequently t .05 dm is 1.671 x .4423 or .739 and t .01 dm is 2.358 x .4423 or 1.04. Similarly, in a one-tailed test, t values of 2.617 x .4423 (1.16) and 3.373 x .4423 (1.49) are required for the .005 and .0005 levels of significance respectively (Popham, 1967, p. 198).

In short, in one-tailed tests, the product of the t values times the standard error of the difference between means requires the differences given below. In each case, the probability of the differences is also given.

Differences	P value
.739 or more	.05 or less
1.04 or more	.01 or less
1.16 or more	.005 or less
1.49 or more	.0005 or less

TABLES

TABLE II

Summary for the analysis of variance on the entire sample over the three blocks

	1964			1968		
Source of Variation	d.f.	M.S.	F	d.f.	M.S.	F
Treatment or Reinforcement	2	203.50	13.01*	2	310.72	23.40*
Subjects	59	156.88	10.03*	59	113.67	8.56*
Interaction or Residual	118	15.64		118	13.28	
Totals	179			179		

* $P < .0005$

TABLE IV.

Significant differences in weighted mean gain scores between blocks of reinforcement for all possible paired comparisons in the entire sample

	1964		1968	
Block Gain Scores Compared	Means Compared	D	Means Compared	D
Block IV. vs. II.	35.78 - 32.13	3.65*	36.383 - 34.467	4.533**
Block IV. vs. III	35.78 - 34.40	1.38*	36.383 - 34.467	1.916**
Block III. vs. II.	34.40 - 32.13	2.27*	34.467 - 31.850	2.617**

* $P < .01$ in a two-tailed test
** $P < .0005$ in a one-tailed test

TABLE V.

Summary on a two way analysis of variance for gain scores by children with four classifications of adjustment, six conditions of reinforcement and interaction

	1964			1968		
Source of Variation	d.f.	M.S.	F	d.f.	M.S.	F
Type of Adjustment (raws)	3	256.33	4.78*	3	63.125	1.3283***
Conditions of Reinforcement or Treatment (columns)	5	194.20	3.50*	5	926.500	3.8990*
Interaction	15	33.87	.60**	15	22.47	0.4727**
Within	156	56.74		144^a	47.52	
Totals	179	61.04		167^a	49.68	

* $p < .005$
** $p > .75$
*** $p > .25$

The df for within groups and totals differ in the two studies because the IBM 360/40 computer did not have a prepared program for a two way analysis of variance with Ns unequal. In the interest of expediting the statistical analysis by avoiding the delay involved in writing a new program, the eighth entry of the 12 groups with eight numbers was eliminated to achieve Ns of equal size. On the basis of consultation with his statisticians, Reiter (1968) gave positive and unequivocal assurance that the results with equal Ns would vary very little from the data of unequal groups.

TABLE VI.

Raw scores for the two way analysis of variance (1964).

Adjustment Classification	Conditions of Reinforcement						Sum and $\overline{X}$ for adj. classification
Male-adjusted	E_1B_2	E_2B_2	E_1B_3	E_2B_3	E_1B_4	E_2B_4[a]	
	24	27	24	34	28	39	
	31	31	39	41	41	43	
	26	27	36	28	42	39	
	45	25	48	31	44	25	
	31	42	34	46	31	47	
	32	45	39	35	40	39	
	42	46	47	48	47	46	
	25		25		25		
Sum	256	243	292	263	298	278	1630
Mean	32	34.71	36.5	37.57	37.25	39.71	36.22
Female-adjusted	E_1B_2	E_2B_2	E_1B_3	E_2B_3	E_1B_4	E_2B_4	
	43	27	34	25	36	26	
	32	25	32	28	34	27	
	36	40	37	43	41	46	
	40	46	42	46	44	48	
	31	44	37	36	41	39	
	28	31	33	28	33	26	
	31	43	30	47	35	47	
		34		38		38	
Sum	241	290	245	291	264	297	1628
Mean	34.43	36.25	35.00	36.37	37.71	37.12	36.18

[a] E_1B_2 = first examiner and second block, etc.

TABLE VI. - Continued

Adjustment Classification	Conditions of Reinforcement						Sum and $\overline{X}$ for adj. classification
Female-maladjusted	E_1B_2	E_2B_2	E_1B_3	E_2B_3	E_1B_4	E_2B_4	
	41	40	34	47	35	45	
	30	47	32	43	34	49	
	29	31	25	32	22	30	
	30	31	36	24	38	28	
	27	35	28	36	26	42	
	20	31	22	34	21	32	
	27	28	25	27	30	27	
		21		33		24	
Sum	204	264	202	276	206	277	1429
Mean	29.14	33.00	28.86	34.50	29.43	34.62	31.76
Male-maladjusted	E_1B_2	E_2B_2	E_1B_3	E_2B_3	E_1B_4	E_2B_4	
	30	32	34	28	39	26	
	27	41	25	42	24	40	
	25	29	25	47	25	48	
	24	25	20	45	27	46	
	29	34	39	43	50	44	
	21	18	25	21	23	43	
	37	28	30	40	31	47	
	30		31		34		
Sum	223	207	229	266	253	274	
Mean	27.87	29.57	28.62	38.00	31.62	39.14	
Sum for cond.	924	1004	968	1096	1021	1126	
$\overline{X}$ for cond.	30.80	33.47	32.27	36.53	34.05	37.53	
Sum for class. of adj.							6139
Mean for class.							34.11

TABLE VII.

Raw scores for the two way analysis of variance (1968).

Adjustment Classification	Conditions of Reinforcement						Sum and $\overline{X}$ for adj. classification
Male-adjusted	E_1B_2	E_2B_2	E_1B_3	E_2B_3	E_1B_4	E_2B_4[a]	
	43	38	47	33	47	42	
	30	27	37	25	37	23	
	40	34	42	43	43	43	
	38	46	40	26	41	26	
	30	34	27	43	32	45	
	26	28	28	34	28	36	
	23	25	38	34	41	38	
	27		31		35		
Sum	257	232	290	238	304	253	1574
Mean	32.125	33.14	36.25	34.00	38.00	36.14	34.98
Female-adjusted	E_1B_2	E_2B_2	E_1B_3	E_2B_3	E_1B_4	E_2B_4	
	25	22	25	25	25	25	
	35	31	35	42	39	42	
	29	22	39	31	38	35	
	31	33	37	39	42	45	
	28	33	36	38	41	36	
	44	30	48	30	48	31	
	37	34	43	33	45	32	
		34		40		40	
Sum	229	239	263	278	278	286	1573
Mean	28.625	29.875	32.875	34.75	39.71	35.75	34.93

[a] E_1B_2 = first examiner and second block, etc.

TABLE VII. - Continued

Adjustment Classification	Conditions of Reinforcement						Sum and $\bar{X}$ for adj. classification
Female- Maladjusted	E_1B_2	E_2B_2	E_1B_3	E_2B_3	E_1B_4	E_2B_4	
	30	29	36	26	43	26	
	34	43	38	38	40	47	
	27	46	25	47	25	49	
	33	30	31	35	30	34	
	24	26	31	22	34	27	
	34	32	45	35	47	36	
	34	28	31	28	30	29	
	29		33		35		
Sum	245	234	270	231	284	248	1512
Mean	30.625	33.43	33.750	33.00	35.50	35.43	33.6
Male maladjusted	E_1B_2	E_2B_2	E_1B_3	E_2B_3	E_1B_4	E_2B_4	
	28	29	44	27	45	32	
	26	28	36	26	36	28	
	41	25	41	25	44	25	
	29	26	26	32	27	34	
	43	32	37	24	45	30	
	39	30	41	34	42	34	
	25	31	26	32	26	34	
		43		47		48	
Sum	231	244	251	247	265	265	1503
Mean	33.0	30.5	35.86	30.875	37.86	33.12	33.4
Sum for cond.	962	949	1074	994	1131	1053	
$\bar{X}$ for cond.	32.06	31.63	35.80	33.13	37.7	35.1	
Sum for class. of adj.							6162
Mean for class.							34.23

3

TABLE VIII.

Means and standard deviations of gain scores

	1964		1968	
	Male		**Male**	
	Mean	SD	Mean	SD
Bl. II.	30.97	7.31	32.13	6.69
Bl. III.	35.00	8.50	34.20	7.20
Bl. IV.	36.77	8.81	36.23	7.37
Total	34.24	8.61	34.25	7.15
	Female		**Female**	
	Mean	SD	Mean	SD
Bl. II.	33.30	7.08	31.57	5.75
Bl. III.	33.80	6.82	34.73	6.72
Bl. IV.	34.80	8.69	36.53	7.46
Total	33.97	7.48	34.28	6.92
	Total		**Total**	
	Mean	SD	Mean	SD
Bl. II.	32.13	7.31	31.85	6.19
Bl. III.	34.40	7.77	34.47	6.91
Bl. IV.	35.78	7.90	36.38	7.36
Total	34.32	7.09	34.24	7.05
	Adjusted		**Adjusted**	
	Mean	SD	Mean	SD
Bl. II.	34.23	7.87	31.90	6.35
Bl. III	36.37	7.14	35.63	6.59
Bl. IV.	37.90	7.23	37.37	6.95
Total	36.20	7.41	34.97	6.95
	Maladjusted		**Maladjusted**	
	Mean	SD	Mean	SD
Bl. II.	29.93	6.49	31.80	6.14
Bl. III.	32.43	7.87	33.30	7.14
Bl. IV.	33.67	9.12	35.40	7.73
Total	32.01	8.04	33.50	7.11

TABLE VIII. - Continued

	1964		1968	
	Male-adjusted		Male-adjusted	
	Mean	SD	Mean	SD
Bl. II	33.27	7.99	32.60	7.05
Bl. III.	37.00	7.75	35.20	6.94
Bl. IV.	38.40	7.32	37.13	7.18
Total	36.22	7.99	34.98	7.15
	Male-Maladjusted		Male-Maladjusted	
	Mean	SD	Mean	SD
Bl. II.	28.67	5.71	31.67	6.52
Bl. III.	33.00	8.76	33.20	7.56
Bl. IV.	35.13	9.85	35.33	7.70
Total	32.27	8.70	33.47	7.36
	Female-Adjusted		Female-Adjusted	
	Mean	SD	Mean	SD
Bl. II.	35.40	6.48	31.20	5.72
Bl. III.	35.73	6.47	36.07	6.44
Bl. IV.	37.40	7.11	37.60	6.96
Total	36.18	7.05	34.96	6.83
	Female-Maladjusted		Female-Maladjusted	
	Mean	SD	Mean	SD
Bl. II.	31.20	6.93	31.93	5.96
Bl. III.	31.87	6.77	33.40	6.95
Bl. IV.	32.20	8.07	35.47	8.03
Total	31.76	7.27	33.60	7.02

TABLE IX.

Comparison of Mean Weighted Gain Scores Between Social Classes in Descending Order of Differences

Classes Compared	D	Sample Size N1	Sample Size N2	Critical U Value Necessary for D ($p < .05$)	U Value Found	Conclusion
III. vs. IV.	6.23	4	13	8 or less	16.5	Accept N.H.
I. vs. IV.	4.30	13	15	54 or less	86.5	Accept N.H.*
III. vs. VI.	3.57	4	7			
V. vs. IV.	3.34	9	13			
III. vs. VII.	3.25	4	4			
VII. vs. IV.	2.98	4	13			
III. vs. V.	2.89	4	9			
VI. vs. IV.	2.66	7	13			
III. vs. I.	1.93	4	15	10.00 or less	25.50	Accept N.H.
I. vs. VI.	1.64	7	15			
I. vs. VII.	1.32	4	15			
I. vs. V.	.96	9	15			
V. vs. VI.	.68	7	9			
V. vs. VII.	.36	4	9			
VII. vs. VI.	.32	4	7			

Means on Gain Scores by Social Classes on Block IV.

Classes	$\overline{X}$	N	Description of Class (U. S. Census, 1963)
I.	38.07	15	Professional and technical workers.
III.	40.00	4	Clerical, sales, and kindred workers.
IV.	33.77	13	Craftsmen, foremen, and kindred workers.
V.	37.11	9	Operatives and kindred workers.
VI.	36.43	7	Service workers.
VII.	36.75	4	Laborers, except farm and mine.

* Null hypothesis of no significant differences

TABLE X.

General summary of significant statistics

Statistic (X)	Comparison Made	Size of (X)		Value of p or interpretation
		1964	1968	
t test	Number of human preferences before and after conditioning (entire sample)	12.29 (58 df)		$p < .001$ (two-tailed)
			12.22	$p < .0005$ (one-tailed test)
t test interpreted as z score	Differences in mean weighted gain scores between adjusted and maladjusted groups on Block 4	1.961 (58 df)		$p < .05$ (two-tailed test)
			1.04	$p > .20$ (one-tailed test)
Chi-square	Relationship between scores above and below the common mean for Block 4 and adjustment status classified on the basis of extreme percentile points on all three scales		(1 df) $X^2=4.20$	$p < .05$ (two-tailed test)
t test	Differences in raw scores between the two adjustment groups on number of human preferences for Block 1 prior to conditioning (baseline scores)	.49 (58 df)	.53	$p > .60$
t test	Sex differences on mean weighted gain scores for Block 4	.89 (58 df)		$p > .20$
			.16	$p > .80$ (two-tailed tests)
t test	Examiner differences in gain scores for Block 4 (entire sample under each E)	1.46 (58 df)	1.40	$p > .10$ (two-tailed tests)

TABLE X. - Continued

Statistic (X)	Comparison Made	Size of (X)		Value of p or interpretation
		1964	1968	
_t_ test	Examiner differences (adjusted pupils)	.65 (28 df)	1.40	$p > .40$ $p > .10$ (two-tailed tests)
_t_ test	Examiner differences (maladjusted pupils)	4.14 (28 df)	.85	$p < .001$ (two-tailed) $p > .20$ (one-tailed)
Chi-square	Examiner differences --relationship between being tested by E_1 or E_2 and number of scores by "culturally deprived" children above or below the common mean for the entire sample on the criterion for Block 4 of the learning task (adjustment status of Ss combined)		X^2 of 5.83 (1 df)	$p < .02$ (two-tailed test)
Pearson r	Correlation between I. Q. scores on the Peabody and gain scores in Block 4 of the learning task (entire sample)	.07	.0653	Negligible relationship
Pearson r	Correlation between socioeconomic scores for occupations and gain scores for Block 4 (entire sample)		.0683	Negligible relationship

TABLE X. - Continued

Statistic (X)	Comparison Made	Size of (X)		Value of p or interpretation
		1964	1968	
X^2	Relationship between type of paternal occupation and scores above and below common mean for Block 4 (white collar and unskilled groups vs. semi-skilled and "service" groups)	6.072 (1 df)		$p < .02$ (two-tailed)
Mann-Whitney U Test	Same as above		z value of .9733 found	p of .1660 (one-tailed test)
Mann-Whitney U Tests	All possible paired comparisons between the gain scores of any two social classes	No statistically significant U values found in either study		
Mann-Whitney U Test	Differences in Prestige Scores between adjustment groups		z score of 4.93 for difference between two adjustment groups	$p < .00003$ (one-tailed test)

TABLE XI.

Summary of differences in weighted gain scores for Block 4 under each examiner within subgroups of each adjustment category (1968)

Adjustment Subcategory Compared	U Value Found	Two-tailed probability of value for $n_1 = 7$ and $n_2 = 8$
Female-adjusted	16	.19
Female-maladjusted	26	.866
Male-adjusted	23	.612
Male-maladjusted	15	.152

TABLE XII.

Relationship between extremes in gain scores and adjustment status (Fisher's exact probability test)[a]

	Maladjusted	Adjusted	Totals
+ 1 S.D. above $\overline{X}$	A. 7	B. 5	12
- 1 S.D. below $\overline{X}$	C. 13	D.0*	13
Total	20	5	25

*P of .05 (two-tailed test)

[a]When the sums of both A & B and C & D are equal to 12, a B value of 5 requires a D value of zero for the .05 level of significance in a two-tail test. Since the above C + D value is equal to 13, a sum which is not listed within the table for critical values in Fisher's test, it appears, at first glance, that the use of significance levels in Fisher's exact probability test is not appropriate here.

When A & B is held constant, however, the effect of an increase in the C + D value is to raise the score of the D value required for any given level of significance. Consequently, if the B value of 5 requires a D score of zero for the .05 level in a two-tail test when C + D equal 12, it is reasonable to conclude that an identical D score under the conditions of C + D equal to 13 will also be statistically significant. Moreover, as Siegel has pointed out (1956, pp. 99) the use of significance levels in Fisher's test errors on the side of increasing the size of the probability. The skeptical reader, however, should consult the detailed table of critical values for Fisher's exact probability test in Siegel's text to draw his own conclusions on the legitimacy of this scoring system (1956, pp. 256-269).